Resource Adequacy Mechanisms:

Spot Energy Markets and Their Alternatives

Scott M. Harvey

Center for Research in Regulated Industries
19th Annual Western Conference
Monterey, California

June 28, 2006
Revised June 21, 2006

Table of Contents

Resource Adequacy Mechanisms: Spot Energy Markets and Their Alternatives[1]

Scott M. Harvey
(Revised June 21, 2006)

I. OVERVIEW

The restructured PJM, New York and ISO-New England electricity markets all include installed capacity as well as energy markets. These capacity markets are the successor to the reserve requirements of the power pools that preceded the ISO operated markets. Over time, these capacity markets have encountered a variety of problems. Some of these problems have been more or less successfully addressed but others have grown more and more intractable, leading to proposals to substantially change the capacity market systems in PJM and New England.

Section II discusses the origins of the current PJM, NYISO and ISO-New England capacity market systems and the problems they were intended to address. Section III turns to a description of the key features of these capacity market systems, organized around eight problem areas: defining capacity requirements; deliverability requirements; retail access; outage performance; unit availability; imports and exports of capacity; demand response and market power. Section IV discusses proposals for resource

---

1 This paper is based upon a paper prepared for the California ISO in the summer of 2005, updating an earlier paper. Portions of that earlier paper also provided the basis for the discussion of Resource Adequacy in Chapter X and Appendix V of Scott M. Harvey, William W. Hogan, and Susan L. Pope, "Comments on the California ISO MRTU LMP Market Design." Earlier versions of this paper benefited from the comments of Jim Bushnell, John Chandley, Steven Greenleaf and William Hogan. The views expressed here are those of the author and not necessarily those of any of the ISOs discussed. The author, along with a number of colleagues, was a consultant to the Member Systems of the New York Power Pool and the PJM Supporting Companies during the development of the PJM and New York LMP-based pricing systems and capacity markets. The author also is or has been a consultant on electricity market design and transmission pricing, market power or generation valuation for Allegheny Energy Global Markets; American Electric Power Service; American National Power, Calpine Corporation; Centerpoint Energy; Commonwealth Edison; Constellation Power Source; Coral Power; Dynegy; Edison Electric Institute; General Electric Capital; Goldman Sachs; GPU; GPU Power Net Pty Ltd; GWF Energy; Independent Energy Producers Association; ISO New England; Longview Power; Midwest ISO; Morgan Stanley Capital Group; New England Power; New York Energy Association; New York ISO; Ontario IMO; PJM; Reliant Energy; San Diego Gas & Electric; Sempra Energy; Mirant/Southern Energy; Texas Utilities; Transpower of New Zealand Ltd; Westbook Power; Williams Energy Group; and Wisconsin Electric Power Company. Any errors are solely the responsibility of the author.

adequacy mechanisms based on call contracts for energy. Section V discusses the potential for resource adequacy mechanisms that blend capacity markets and energy shortage pricing. Section VI summarizes the conclusions.

II. WHY CAPACITY MARKETS?

A. Origins of Capacity Markets

Under the traditional vertically integrated utility model, resource adequacy standards were resolved between the individual utility and its regulators. The consequences of inadequate utility resources to meet utility load were straightforward, the utility that lacked sufficient generation to meet its load needed to buy energy and schedule transmission to import additional power or it would have to undertake involuntary load shedding. The determination of which LSE would shed load during shortage conditions was easy; this was the utility that was short of power or did not have firm transmission service to deliver power to its load. The need for resource adequacy mechanisms, such as installed reserve requirements, the precursor of capacity market systems, initially arose in the Northeast from the implementation of economic dispatch which eliminated the link between an individual utility's generation and load. Individual utilities bought and sold power through the pool and their generating units followed pool dispatch instructions. An individual utility might be a net buyer during a shortage not because it was short of capacity, but merely because that utility's generation was the lowest cost source of operating reserves or regulation. Moreover, the pool dispatch was based on administrative costs rather than market-based rates and intra-pool settlements were typically based on split savings systems rather than market prices, so intra-pool settlement prices were low even during shortage conditions. This operating environment led to rules providing for shared responsibility for load shedding within the impacted region of the pools, rather than attempting to assign responsibility for load shedding to the generation-short distribution company.[2]

Maintaining the capacity needed to meet peak load on a one-day-in-ten-year reliability criteria is very expensive on a per MWh basis, however. Moreover, because marginal capacity will almost never be used, maintaining this capacity can materially raise the overall cost of meeting load. Shared responsibility for load shedding and cost-based pricing therefore gave rise to the prospect that individual utilities would choose to reduce their costs by not incurring the high cost of maintaining the capacity needed to meet their peak load at conventional reliability levels, knowing that energy purchases would be settled at "cost" based rates and most of any resulting load shedding would be

---

2 Of course, to the extent that only a single distribution company served load within the constrained region in which there were inadequate resources available to meet firm load, the load shedding would fall entirely on the responsible distribution company. This will not necessarily be the case, however.

borne by the customers of other utilities. Installed reserve requirements, the pre-cursor of capacity markets, therefore arose in part to ensure that all pool members incurred the cost of maintaining the capacity needed to meet peak day load on a reliable basis in an environment of cost-based pricing and shared responsibility for load shedding,.

A similar logic may have operated in the Midwest under some of the reserve sharing agreements. If the members of a reserve sharing group are to agree to shared activation of reserves, there is a parallel need to make sure that all of the entities benefiting from the reserve sharing agreement incur the cost of maintaining enough capacity to provide reserves to the other members of the reserve sharing group under stressed system conditions. The MAPP region in particular appears to have developed rules designed to avoid free riding under the reserve sharing agreement and ensuring that all participants bore the costs of maintaining capacity adequacy.

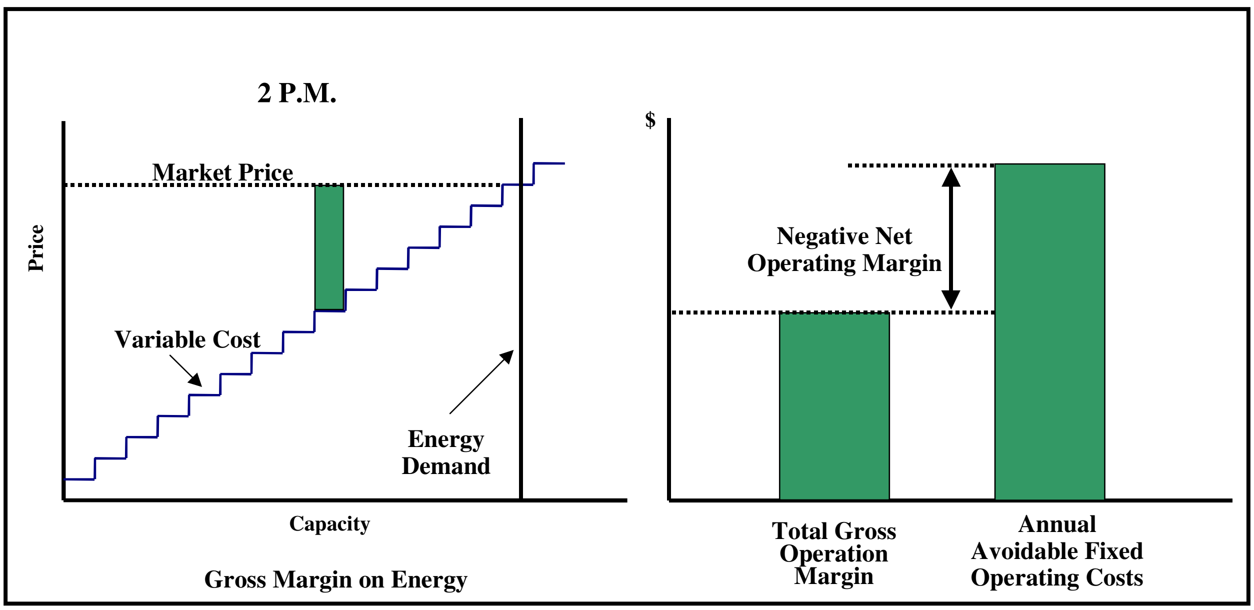
Importantly, these reliability structures could not rely on prices to allocate energy within the pool or reserve sharing group during shortages as energy was bought and sold at cost based rates that did not reflect the value of energy or capacity during these shortage conditions.

One alternative for maintaining reliability within ISO coordinated markets of the Northeast pools when the pools transitioned to ISO dispatched open access markets was therefore to maintain the reserve requirements of the power pools in some form as a reliability mechanism. The need for such a reliability mechanism was increased by the $1,000/MWh bid cap, the imperfect shortage pricing that existed at start up of the PJM and NYISO capacity markets, and the intent of several states to utilize transmission open access to support retail access programs.

B. Capacity Market Systems

Price equal to the short-run marginal cost of the marginal supplier is a basic short-run equilibrium condition. With the introduction of market-based marginal cost pricing in energy markets, infra-marginal generation earned revenues on sales of energy and ancillary services, earning margins equal to the difference between its revenues and the variable costs incurred in generating energy, as portrayed in Figure 1. Generation also incurs fixed costs, some of which can be avoided if the generation owner chooses not to make its capacity available for operation (i.e., if the capacity is either mothballed or closed permanently). In the absence of a capacity or installed reserve requirement, generation owners will not choose to keep capacity in operation for dispatch unless their gross operating margin exceeds their avoidable fixed operating costs.

Figure 1
Generator Operating Margins



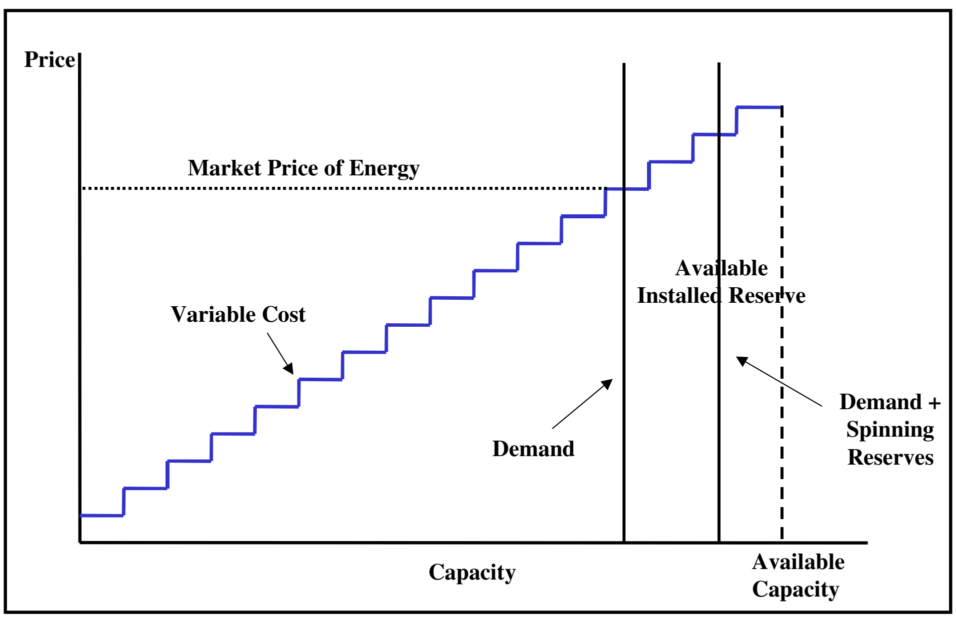
Under a capacity market system, the capacity requirement is established such that reserve shortages are expected during a small number of hours on an annual basis and there is almost always sufficient capacity to avoid involuntary load shedding.[3] The wholesale energy market therefore usually clears at the intersection of demand and the

3 The expected number of hours of reserve shortage (i.e., emergency state operation during the year) is relatively low but the frequency of reserve shortage exceeds the one-day-in-ten-year load shedding standard. Only severe reserve shortages result in load shedding. As discussed below, it is costly and undesirable to operate in an emergency state, so the desired level of capacity should be established, recognizing the costs to consumers of an emergency state in addition to the costs of load shedding when it occurs.

variable cost (dispatch) curve, as portrayed in Figure 2. Because the price of energy is generally set by the incremental cost of the energy generated by marginal units, the price is not high enough often enough to cover the full cost of keeping these marginal units in operation over the year (i.e., the units will have a negative net operating margin as portrayed in Figure 1), unless the energy price is extremely high during hours of reserve shortage.

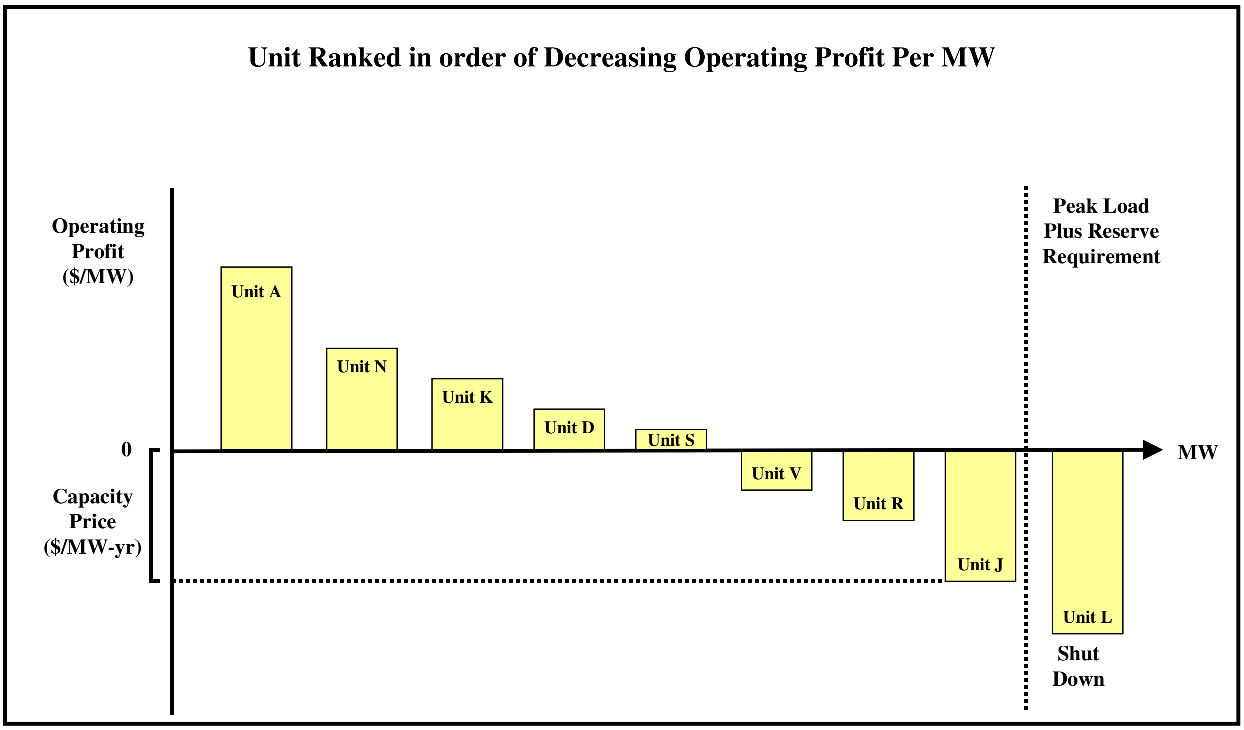
Absent some form of shortage pricing during reserve shortage conditions, many resources needed to maintain reliability will have a negative net operating margin. Under a capacity market system, a market-wide capacity requirement is imposed symmetrically on all load-serving entities within the market. LSEs may not have the traditional obligation to serve, but under a capacity market system they must demonstrate resources sufficient to meet the installed capacity requirement for their customers. If the amount of generation required to be available under the installed capacity requirement exceeds the amount of generation that would have been available in the absence of such a requirement (i.e., the amount justified by energy and reserve market revenues alone), a market for capacity is created. Thus, with such a binding capacity market requirement, capacity takes on value in and of itself. Marginal units, unprofitable on the margins they earn on energy sales and ancillary services, would demand a capacity payment in return for agreeing to make themselves available for operation and allowing the contracting LSE to satisfy its capacity market requirement.

Figure 2
Energy Market Prices with Installed Capacity Requirements



To keep capacity open under a capacity market system, the owner of the marginal resource requires a capacity payment of at least the difference between its avoidable fixed operating costs and its net margin on energy and ancillary services sales (i.e., it must recover the negative net operating margin portrayed in Figure 1 on an expected value basis). Competition among capacity owners and with potential entrants should cause the market-clearing capacity payment to approximate the per-MW payment that would induce just enough generation to remain available to enable the capacity requirement to be met. Under a market-based capacity system, all generating resources contracting to provide installed reserves are paid the market-clearing price of capacity, as portrayed in Figure 3. Between the capacity payments they receive and their margins on energy sales and ancillary services, all resources providing capacity needed to meet the capacity requirement would earn enough to cover their avoidable fixed operating costs and thus would remain available.

Figure 3
Determination of Market Price of Capacity



With a capacity market requirement, the capacity payment is determined by the per-MW payment required to enable the marginal resource (Unit J in Figure 3) to at least break even and capped by the payment required to keep the next most expensive resource in operation (Unit L). For existing capacity, the breakeven point would be based on going-forward costs while for a new entrant the breakeven point would include a return of and on investment. Because the market can meet the capacity requirement without Unit L, the market-clearing capacity payment would be insufficient for it to cover its anticipated operating losses and Unit L would close. Between the capacity payments they receive and their margins on energy and ancillary services sales, each of the other units remaining open would make at least enough to cover their avoidable fixed operating costs. Because capacity market suppliers would be paid the market-clearing price, most incumbent capacity market suppliers would earn more than their going-forward costs. This is inherent in a market-based system. Conversely, however, incumbent suppliers would not be assured of returning a return of and on their investment as under rate-of-return regulation, but only of recovering their going-forward costs.

Installed capacity systems have several potential limitations:

- A capacity market system ensures that the electricity market clears by keeping generating capacity that cannot recover its costs in the energy and ancillary services market in operation. The cost of keeping this capacity available may exceed its actual value to consumers.

- A set of rules is required to govern the location of qualifying capacity.

- A set of rules is required to govern generator operational availability.

- A set of rules is required to govern the treatment of imports.

- There is a potential for free-riding by any loads not required to maintain installed capacity.

- Low energy prices during shortage conditions mean that there will be too little incentive for loads to become price-responsive in real time unless this incentive is built into the capacity market system.

- Absent additional rules, a capacity market system ensures the availability of capacity but does not ensure that energy is available in any particular quantity at any particular price from this capacity.

- There is a potential for a short-term capacity market system to become little more than a second payment for energy.

- There is a potential for the exercise of market power that can be difficult to address without undermining other policy goals (reliability, retail access).

These issues are discussed in Section III. Before turning to a discussion of these issues, it will be helpful to first discuss the alternative of relying solely on energy and reserve pricing to maintain resource adequacy and reliability.

C. Energy-Only Pricing

An alternative to capacity market systems in maintaining resource adequacy is to structure energy and ancillary service markets such that the marginal generator is able to recover its going-forward costs in energy and ancillary service prices. In principle, the market design elements needed to sustain such an energy market-based resource adequacy system are to implement shortage pricing that causes the prices of energy and ancillary services to rise to a sufficiently high level during reserve shortage conditions that the marginal capacity resource required to meet established reliability criteria is able to recover its going-forward costs during these reserve shortage hours.[4]

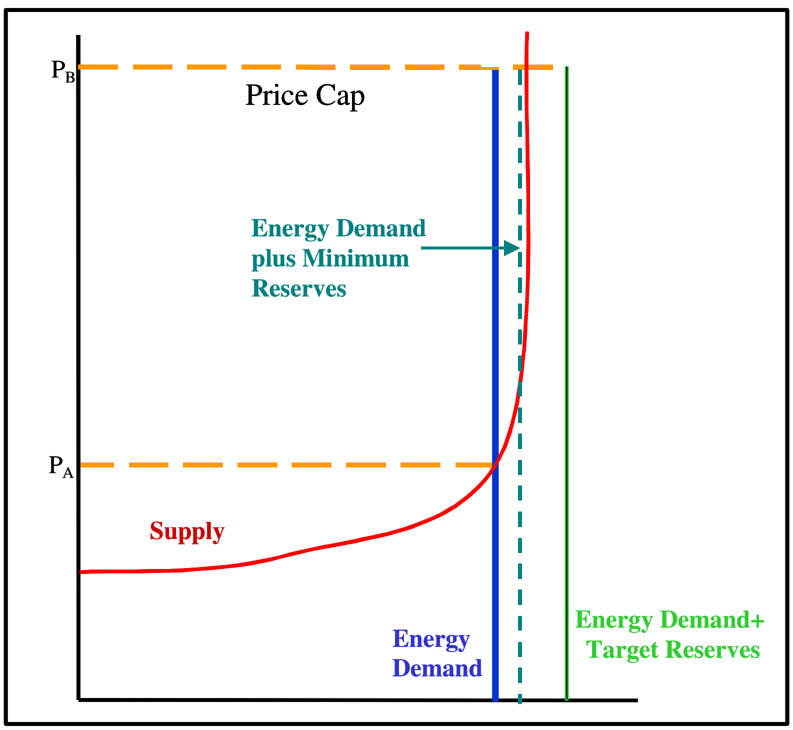
For vertically integrated utilities such an energy only market could operate much like a capacity market system. The pool operator/ISO could determine the shortage prices that it estimates are required to keep sufficient capacity available to meet the target level of reliability and could inform the vertically integrated utilities of the implied reserve margin. The shortage pricing would support the implied reserve margin as the pricing system would be designed such that there would be enough hours with high prices to justify keeping the target level of capacity in operation. Nevertheless, there are some reliability risks in this market design and these risks are magnified in markets with unintegrated retailers and suppliers.

---

[4] It is sometimes suggested that consumer demand that is highly or even complete price-inelastic in the short run makes it difficult or impossible to rely on energy-only markets to maintain reliability. (See, for example, Juan Rosellon, "Different Approaches to Supply Adequacy in Electricity Markets," and Peter Cramton and Steven Stoft, "The Convergence of Market Designs for Adequate Generating Capacity," April 25, 2006 (hereafter Cramton and Stoft 2006).) Highly price-inelastic retail power demand complicates the problem of maintaining reliability whether reliability is maintained through the decisions of a vertically integrated utility or a market-based generation supply mechanism but does not necessarily favor a particular reliability mechanism.

A fundamental feature of an energy-only market design based on shortage pricing is that with vertical demand and competitive markets, the market price of energy and reserves only exceeds the incremental costs of the marginal generator when the control area is reserve short. As long as these reserve shortages are small, they will have little impact on reliability, so conventional reliability standards can be consistent with an energy shortage pricing system. Thus, as portrayed in Figure 4, the price would rise to the price cap when reserves fell below the target level, but involuntary load shedding would occur only when capacity fell below a lower threshold, labeled energy demand plus minimum reserves in Figure 4.[5] A practical difficulty in implementing such a market design is that actual peak load is uncertain, as is the available capacity (due to random outages). In consequence, the more often the system is expected to be in a reserve short condition given normal weather and outages, the greater the potential for bad luck in terms of weather or outages to throw the system into the range in which involuntary load shedding is required.

Figure 4
Supply and Demand in a Shortage



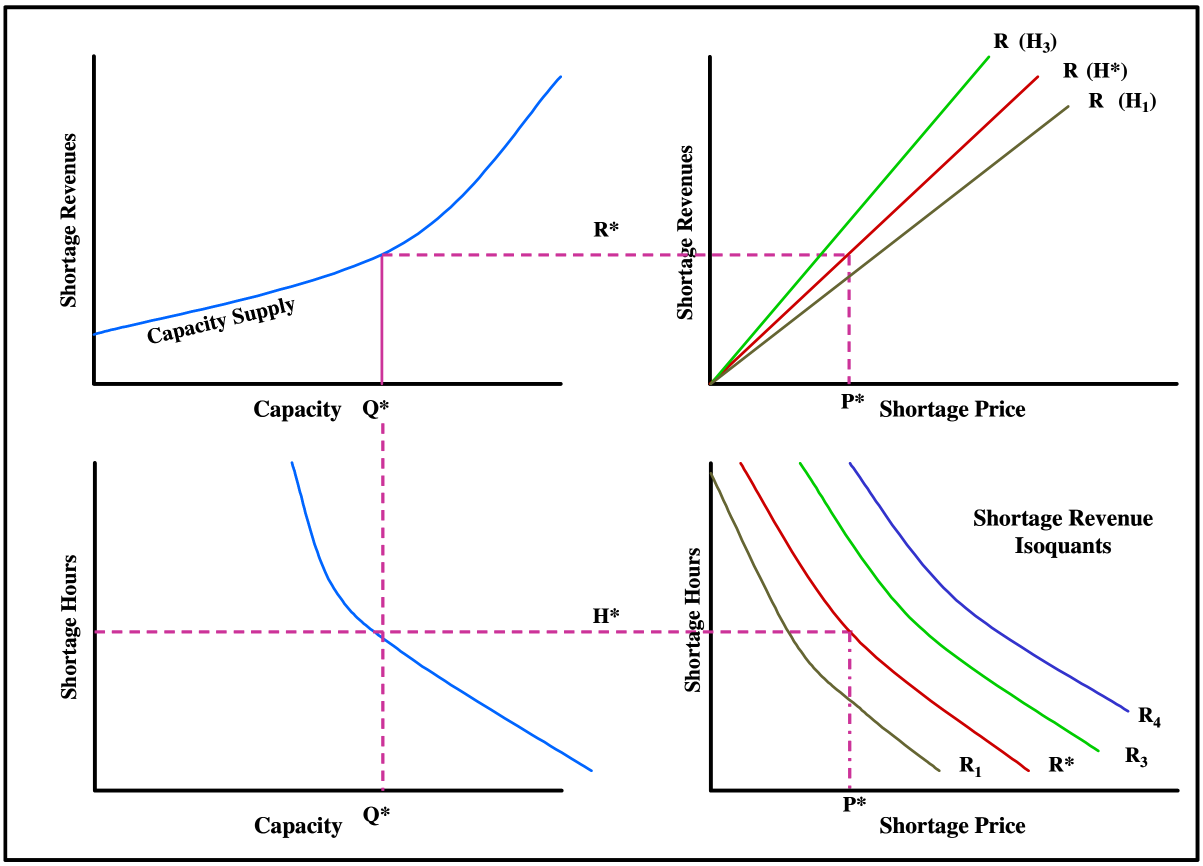
---

5 Cramton and Stoft describe operating reserve requirements and load shedding reserve thresholds as market design choices but these security criteria are determined almost entirely independent of economic criteria and are not market design choices. Cramton and Stoft 2006, pp. 4, 9, 20-21, 26-28.

If the short-term demand for electric energy is completely price-inelastic, the capacity needed to maintain a given level of reliability is the same under an energy-only pricing system or under a conventional reserve margin or capacity market system. This follows tautologically from the assumption that demand is completely price-inelastic, since the peak load is not affected by the pricing system. The potential complexity under an energy-only pricing system is that a shortage pricing system must be implemented in such a manner that the capacity needed to maintain the desired reliability level earns its going-forward costs during the number of hours of reserve shortage that are consistent with the desired level of reliability.

An important observation is that there is always a shortage price that satisfies the equilibrium condition, even if demand is completely inelastic. The left-upper hand panel of Figure 5 portrays the supply and demand for capacity. The supply of capacity is upward sloping, with higher-per-MW shortage revenues required to sustain higher levels of capacity. There is some level of shortage revenues (R*) that will sustain a target level of capacity Q*. Q*, in turn, implies some expected number of shortage hours (H*) and the right-hand panel shows that, for any given number of shortage hours, there is some shortage price (P*) that results in the level of shortage revenues required to sustain Q* capacity.

Figure 5
Capacity, Shortage Hours and Shortage Prices



While an energy-only pricing system can sustain reliability with price-inelastic demand, there are four basic reliability risks in relying on an energy-only pricing system to meet price-inelastic demand:

- Miscalculation of the cost of capacity by the pool/ISO, resulting in too little capacity in operation to maintain intended reliability levels.

- Miscalculation of expected prices by LSEs/suppliers, resulting in too little capacity in operation to maintain intended reliability levels.
- The reserve shortage frequency required to sustain the marginal unit is inconsistent with intended reliability levels.
- Mismatch between energy and reserve pricing and reliability requirements.

Each of these risks is discussed below.

Miscalculation of Capacity Costs

Under a capacity market system the regulator and/or system operator determines the capacity requirement through Monte Carlo type analysis of reliability under stressed system conditions. Importantly, the calculated capacity requirement does not depend on the cost of having capacity available during stressed system conditions. Instead, the system operator determines the physical capacity required to maintain the target level of reliability. The cost of keeping this capacity in operation is determined in the market through the supply decisions of resource suppliers.[6]

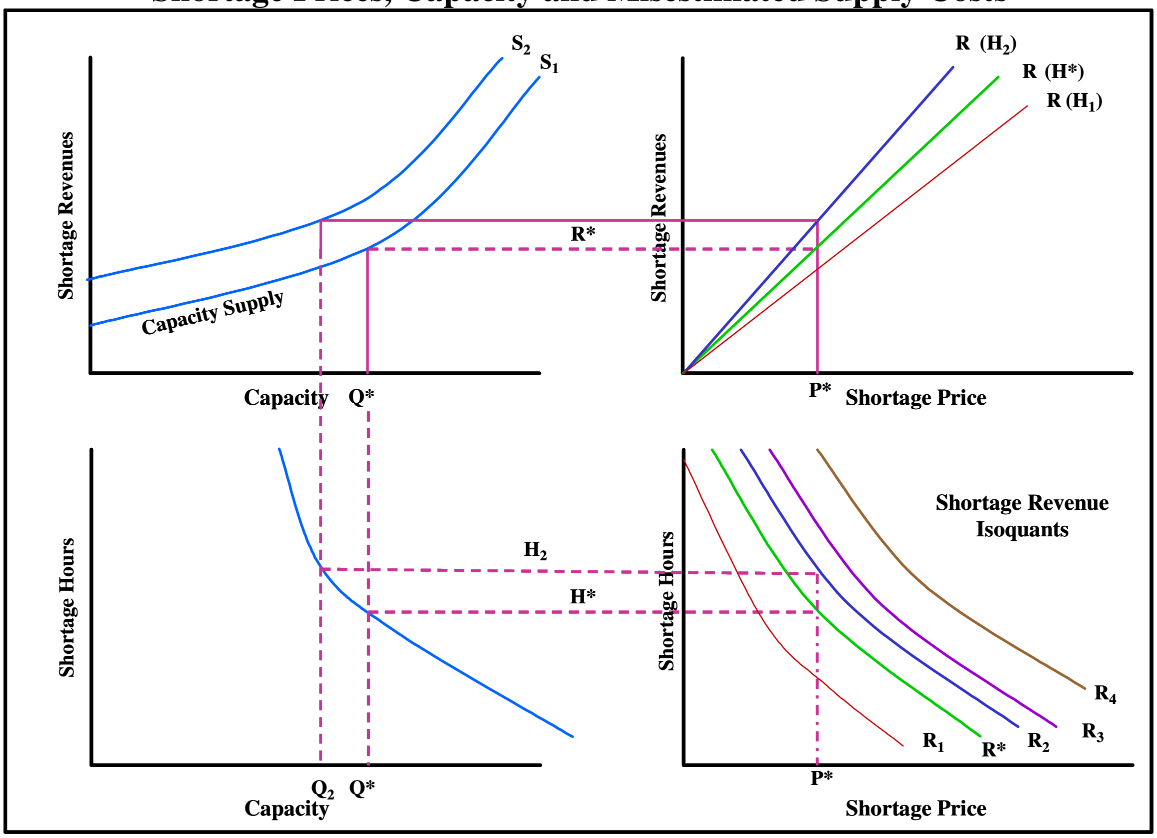
Under an energy pricing system driven by shortage pricing, however, the amount of capacity that will be made available by resource suppliers in response to any set of shortage prices depends on the cost of supplying this level of capacity during shortage conditions. If the system operator misunderstands the cost of having capacity available or miscalculates the revenues generated by marginal capacity during non-stressed conditions, then a given set of shortage prices may result in more or less capacity being available than expected by the system operator, potentially resulting in a different level of reliability than planned for by the system operator and regulators. Since the system operator does not participate in commercial markets there is a potential under energy-only pricing with price-inelastic demand for the system operator to significantly misassess the cost of supplying generating capacity during peak conditions, resulting in more or less capacity being available than assumed in the system operator's reliability analyses.

---

6 If the load-serving entities are vertically integrated utilities, then the cost of meeting the capacity requirement is determined by their cost of building or contracting for the required level of capacity.

This is illustrated in Figures 5 and 6. If the system operator believes the supply curve for capacity to be S_1, it would infer that shortage revenues R* are required to sustain Q* of capacity, which would, in turn, imply a shortage price of P*, as illustrated in Figure 5. Suppose, however, that the actual supply curve is S_2 as portrayed in Figure 6. Then shortage revenues R* will cause suppliers to provide much less than Q* of capacity and the level of capacity will not be sufficient to provide the intended level of reliability.

Figure 6
Shortage Prices, Capacity and Misestimated Supply Costs



If there is a strong link between the shortage costs used by the system operator to determine real-time prices and the actual reliability value of capacity during those conditions, errors in assessing the cost of capacity might not be important in terms of its impact on consumer welfare as the shortage prices would reflect the value of the capacity and should result in an efficient level of reliability, albeit possibly not the expected level of reliability. In a transition between a capacity-based resource adequacy mechanism and a price-based resource adequacy mechanism, however, there is a potential for misunderstanding of how the energy market design will operate in practice to result in a different than intended effective level of shortage prices. Absent such a link between shortage prices and the value of reliability, there is a potential for misestimation of capacity costs by the system operator to lead to a material difference between the actual and intended level of reliability.

These risks are fundamentally transition risks between a capacity requirement based resource adequacy mechanism and an energy price based resource adequacy mechanism. There is a potential for unintended changes in the level of reliability as a result of unanticipated consequences of changes in market design. The converse risk exists in switching from an energy price based resource adequacy mechanism to a capacity requirement based resource mechanism. In much of the MISO footprint, MAIN and ECAR in particular, historically there was not an effective capacity requirement system and resource adequacy was driven by the cost of buying power during shortage conditions, such as those prevailing during 1998. In the MISO the opposite risk exists, moving from a price based resource adequacy mechanism to capacity based resource adequacy mechanisms may adversely impact reliability because of a failure to fully reflect the incentives created by the price based resource adequacy mechanism in the capacity market design or may adversely impact consumer costs because of a misunderstanding of how the capacity market design would operate in practice, resulting in a much different than intended level of capacity prices.

Miscalculation of Expected Prices

While suppliers have a good sense of the overall cost of keeping their capacity available, both suppliers and LSEs may have considerable difficulty projecting expected annual net revenues based on the system operators' shortage pricing rules. The expected price level depends on generation outage probabilities of all resources, not just the suppliers, transmission outages and the supply of imports, as well as the shortage pricing rules. If suppliers and LSEs have different expectations than the pool operator about the frequency and degree of shortage conditions, then they will not provide the anticipated level of capacity in response to a given set of shortage prices, even if the system operator accurately assesses the cost of providing this capacity. One way of potentially achieving consistent expectations across the system operator, suppliers, LSEs and regulators would be for the system operator to make public its profile of simulated shortage prices. While the availability of the system operator's assessment would likely help somewhat converge expectations, market participants would not necessarily find it commercially reasonable to rely on these forecasts.

In principle, market participants should over time be able to assess the accuracy of the system operator price forecast or develop their own expectations, but the reality is that forecasts by resource suppliers and the system operator will be based on expected conditions. Even an accurate forecast may only average out to reflect actual prices over a period of a number of years, and expected conditions may be changing more rapidly than the average of actual outcomes converges on forecasts. It may therefore be difficult for market participants to distinguish whether price estimates are biased ex ante or are accurate estimates of volatile conditions and prices.

Similarly, under a capacity system, the actual level of reliability would differ from the intended level of reliability if the system operator held incorrect expectations regarding generation or transmission outage probabilities as the nominal capacity reflected in the capacity requirement would not produce the level of available capacity assumed by the system operator. Under an energy-only pricing system, such mistaken assumptions regarding outage probability by the system operator would impact the actual level of reliability to the extent they were shared by generators and LSEs. If the suppliers and LSEs had information regarding the probability distribution of aggregate level of outages that is superior to the system operator's information, the decentralized decision making under an energy-only market design could potentially lead to a better assessment of capacity needs than under a capacity market system.

Another potential source of divergent expectations regarding future spot prices under energy-only pricing is the assessment by resource suppliers of the likelihood that regulators will permit market prices to be used for settlements during shortage conditions. Thus, if the reserve shortage price and nominal price cap were $10,000/MWh, the system operators' assessment might be that the marginal supplier would recover its entire net operating revenue shortfall of $50,000/MW during eight hours of shortage conditions in which the price of power would exceed $5,000/MWh. If resource owners did not believe that they would ultimately be permitted to earn more than $1,000/MWh during shortage conditions, the revenue assessment of the system operator and market participant would be radically different, even with common expectations regarding outage probabilities. In this situation, much less capacity might be supplied at a given reserve shortage price than assumed by the system operator.

Resource suppliers are, in principle, able to observe regulators' willingness to allow high prices, providing a feedback loop. There are, however, a few complicating factors in practice. First, the uncertainties involving weather, outages and the number of shortage hours generally will not average out over the course of a single year. Rather, for given supply and demand conditions with a given expected number of shortage hours, the actual number of shortage hours will vary from year to year, with relatively few shortage hours in some years and more in other years.[7] Absent hedging contracts, Supplier cost recovery will be concentrated in years with above-average number of reserve shortage hours; these years will also have the highest costs for consumers and give rise to the greatest pressure for regulatory intervention to reduce consumer costs. The observation by market participants that regulators allow prices to reach shortage levels during years

---

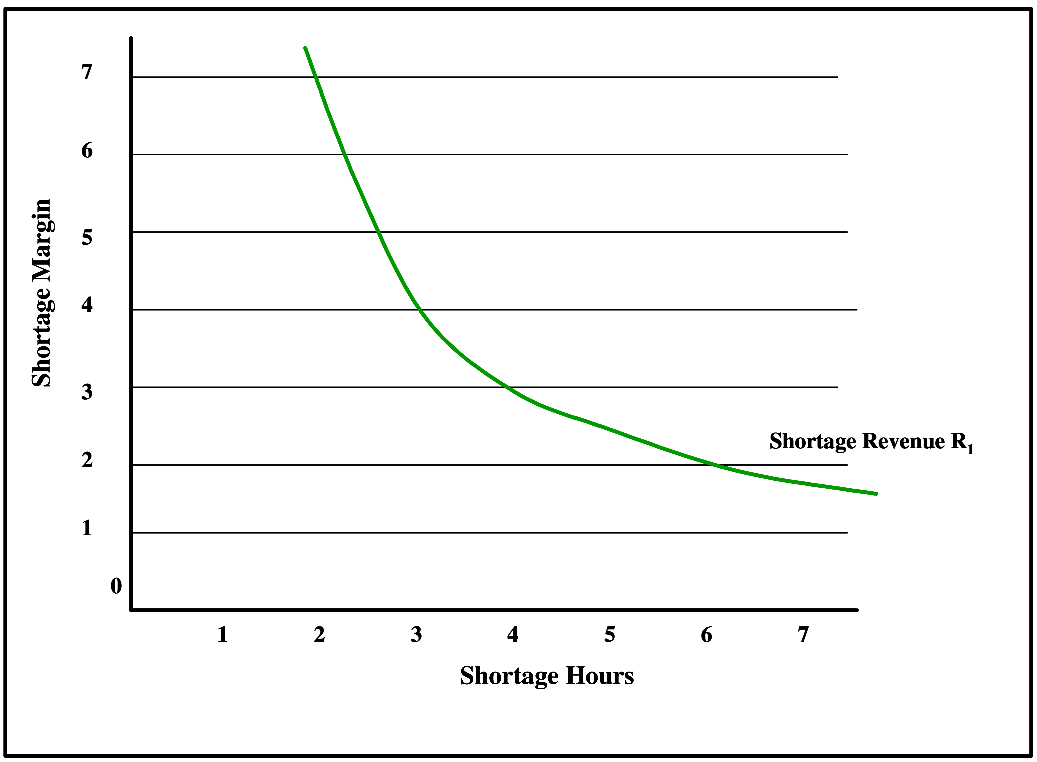
7 This regulatory risk would be most extremely if fixed cost recovery were limited periods of load shedding resulting from inadequate supply in the wholesale market (as opposed to load shedding due to problems on the distribution system) that occurred on average only every ten years, or perhaps less often. In practice, however, reliability not only involves avoiding load shedding but also avoiding the need to enter an emergency state (reserve shortages) which will occur more often than every ten years. Even reserve shortages are not spread evenly year to year so fixed cost recovery would still be concentrated in particular years.

with relatively few shortage hours may not persuade suppliers and LSEs that regulators would allow this outcome in a year with many shortage hours.

Required Shortage Frequency

Under an energy-only pricing system, there is a very explicit tradeoff between the expected price level during shortages and the number of shortage hours required to recover a given net operating cost shortfall. As illustrated in Figure 7, the higher the reserve shortage price and margin, the fewer hours of reserve shortage are required for the marginal resource to recover its going-forward costs. Lower reserve shortage prices therefore imply lower equilibrium levels of available capacity, resulting in a correspondingly higher number of reserve shortage hours in which the going-forward costs of the marginal supplier would be recovered.

Figure 7
Shortage Prices and Shortage Hours



Lower shortage prices and lower capacity levels have reliability consequences, however, as the greater the number of hours of reserve shortage, the greater the likelihood that the reserve shortages will be sufficiently severe in some hours to require involuntary load shedding. Thus, the lower the shortage price in the energy market, the larger the number of shortage hours required for suppliers to recover a given operating cost shortfall in energy prices, and the more likely that load shedding will be necessary during some of the shortage hours.

As suggested in the illustration above, with a price cap of $10,000/MWh and effective shortage pricing, a small number of shortage hours in which the system operator would buy energy offered at the price cap in order to maintain reserves[8] could be sufficient for the marginal generator with an incremental running costs of $100/MWh to recover $50,000/MW in going-forward costs. Given this small number of shortage hours, the probability distribution of demand and supply surprises might, for example, yield a one-day-in-ten years probability of such a large capacity shortage that load shedding was required.

[8] This description assumes that market participants are permitted to submit "hockey stick" bid curves on which the last MW below their ramp rate would be offered at the bid cap so that when this MW was dispatched in order to restore reserves on other units, prices would rise to the price cap. Suppose, for example, that a 500 MW resource has a 5 MW per minute ramp rate and offers 449 MW at $55/MWh and the last 51 MW at $1,000/MWh. The system operator could carry 50 MW of reserves on the resource and generate 449 MW of energy at a price of $55/MWh, and 1 MW would not be dispatched for energy nor provide 10-minute reserves. In a reserve shortage situation, the system operator would be required by reliability criteria to dispatch the 1 MW at $1,000 in order to back down generation elsewhere to restore reserves. This form of hockey stick bidding would not reflect the exercise of market power and would have little impact on economic efficiency if their source submitting the hockey stick bid were the lowest-cost source of contingency reserves. If the resource submitting the "hockey-stick" bid would be a low-cost source of energy relative to other generators, these hockey stick offers could understate the opportunity cost of carrying reserves on the resource and fail to maximize economic efficiency. For this reason, implementation of explicit reserve shortage pricing would, in general, be a more efficient mechanism for reflecting reserve shortage conditions in energy prices. If supplier offer prices are subject to mitigation that holds them below the price cap during shortage conditions, an explicit shortage pricing mechanism would be necessary in order to maintain reliability. See Scott M. Harvey and William W. Hogan, "Issues in the Analysis of Market Power in California," October 27, 2000, pp. 16-25.

Suppose, on the other hand, that the price cap were $1,000/MWh and there was no other form of shortage pricing. The most that the marginal generator could recover during a shortage hour would be $900/MWh. The number of shortage hours required for the marginal generator to recover its going forward costs on an expected basis would be around 55 hours per year. A capacity balance tight enough to produce 55 hours per year of shortage conditions, however, would likely have a much greater risk of requiring load shedding than if only 8 hours were expected, and the increase in likelihood might be non-linear.

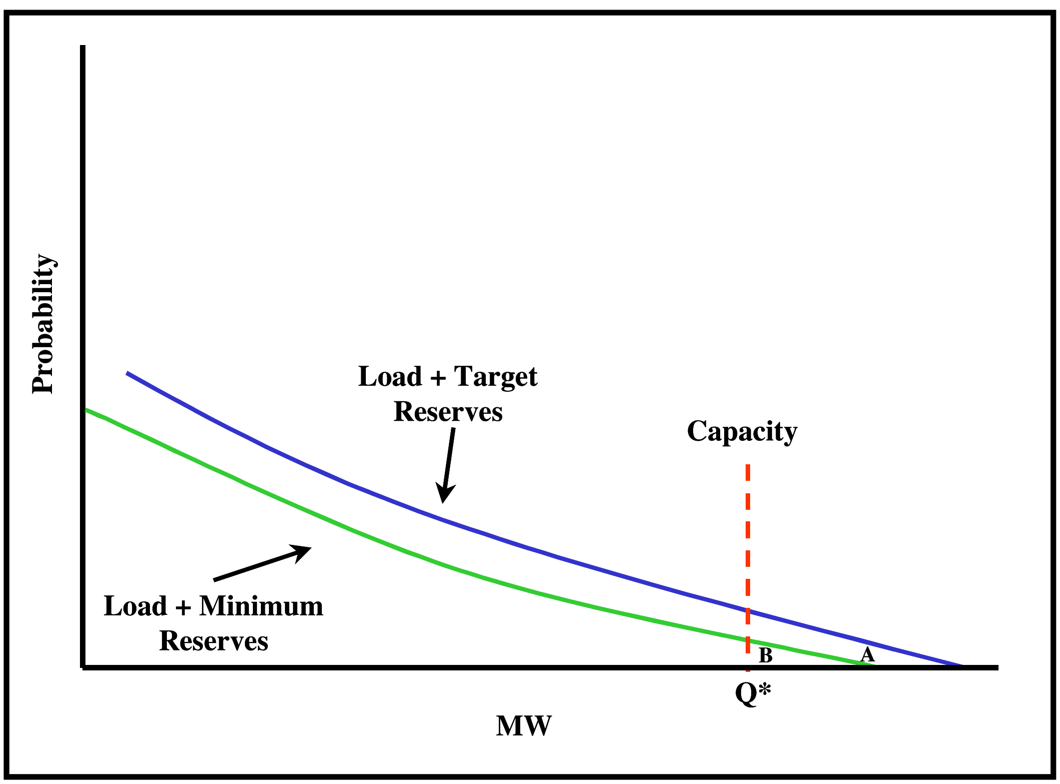
At the extreme, suppose the price cap was set at $250/MW with no other shortage pricing as in California. In this circumstance around 333 hours of reserve shortage would need to be expected on an average annual basis for such a marginal generator to recover its going forward costs in energy prices alone. Such a high frequency of reserve shortages would in turn produce a very high probability of involuntary load shedding.[9]

---

[9] This discussion focuses on the relationship between the expected number of reserve shortage hours and the probability of load shedding. It needs to be kept in mind that it is also costly to consumers for the power system to enter an emergency state as a result of reserve shortages even if no load shedding results. The system operator is required to take a variety of actions to restore reserves that include interrupting non-firm customers and voltage reductions. Moreover, if the system enters an emergency state the operators of the affected transmission and distribution systems must prepare to implement load shedding and consumers must conduct their activities in recognition of the possibility that power could be lost with little or no notice.

Figure 8 portrays this dilemma in terms of the probability distribution of load plus the target level of reserves, and load plus the minimum level of reserves. In this illustration, a shortage arises and prices rise to the price cap when available capacity is less than load plus the target level of reserves. In Figure 8 the probability of this occurring for capacity level Q* is the region A+B. Since lower capacity levels imply larger areas A+B, in a competitive market in long-run equilibrium, capacity would enter or exit until the expected number of shortage hours was large enough given the price cap (or shortage pricing provisions) to make it profitable to keep the remaining capacity in operation. One tradeoff is that the lower the level of capacity, the larger is area B, which is the probability of involuntary load shedding. If the difference between the minimum level of reserves and the target level of reserves is small relative to the slope of the probability function of loads, then high shortage pricing levels will be required for the profits earned during the shortage conditions that occur with probability A+B to support the level of capacity required for probability B to be appropriately small. Moreover, the smaller the difference between the minimum level of reserves and the target level of reserves relative to the slope of the probability fraction of loads, the greater the potential for mistaken evaluation of the probability A+B to result in a load shedding probability (B) that differs from the efficient level.

Figure 8
Capacity, Reserve Shortages and Load Shedding



Low reserve shortage prices that imply a large number of reserve shortage hours on an expected annual basis are therefore unlikely to be efficient for two reasons. First, large numbers of reserves shortage hours will very likely imply an inefficiently large number of hours in which involuntary load shedding is required. Second, a large number of reserve shortage hours would likely itself be inefficient in terms of the costs imposed on consumers by the need to operate in an emergency state for so many hours.

An energy-only pricing system is therefore likely consistent with conventional reliability levels for price-inelastic demand only if energy and reserve prices are very high during shortage conditions so that resource suppliers are able to recover their going-forward costs during a relatively small number of reserve shortage hours. This requirement can be problematic from two perspectives. First, if even small reserve shortages result in prices of $5,000/MWh or more, there would potentially be an incentive for energy and/or reserve suppliers to physically withhold capacity in order to produce artificial reserve shortages and drive energy and reserve prices to the high levels associated with shortage conditions. Second, the year-to-year variability of supply and demand conditions will make it likely that the number of reserve shortage hours will not be relatively constant from year to year. As a result, resource suppliers will not recover their going-forward costs evenly year to year; the recovery will instead likely be concentrated in particular years.[10] Absent forward energy contracts hedging consumers against variations in spot prices, this distribution of reserve shortage hours over time might entail large annual variations in retail prices that may not be practical within the regulatory structure, with the consequence that these large price variations under an energy-only pricing system would magnify regulatory risks.[11]

The continued reliance on capacity markets after several years' experience with deregulated generation markets appears to be motivated in large part by a perception that a capacity market system avoids the potential for the exercise of market power that can exist under energy-only pricing systems and that it avoids the substantial price volatility that will likely exist absent forward contracts under energy-only pricing systems that permit energy and reserve prices to reach sufficiently high levels during shortage conditions to enable resource suppliers to recover their going-forward costs. As discussed below, neither of these perceptions is necessarily valid.

The potential variability of consumer costs from year to year can potentially be addressed through forward energy contracts that hedge consumers against variations in the price of energy. The limitation of this resolution is that unless both LSEs and

---

10 Thus, a marginal generator with a going forward cost of $50,000/MW year might anticipate recovering $15,000/MW year in most years but recovering $200,000/MW year every five years or so.

11 This price pattern could sustain multi-year energy contracts that would recover generator going-forward costs. Under retail access systems, however, there may not be many multi-year energy contracts and thus most customers would be exposed to energy prices during the one year in five or six in which suppliers recover their going forward costs.

suppliers expect effective shortage pricing that will result prices high enough to support the target level of capacity, forward contract prices will not resolve the resource adequacy problem. Forward contract prices will reflect the level of expected future spot prices and if there is no effective shortage pricing in the market design or if there is an expectation that the market design would be changed were high prices to result, then forward contract prices will mirror low spot prices and not support either the entry of new capacity nor the continued operation of high cost existing capacity needed to maintain conventional levels of reliability. The latter problem is particularly acute in U.S. power markets because after FERC's actions to retroactively reduce spot prices in western energy markets during the shortage conditions that prevailed during 2000 and 2001, pricing systems based on the presumption that high spot prices will prevail during shortage conditions will have ever limited credibility. If LSEs do not believe that FERC will actually allow high prices to prevail during shortage conditions, they will not enter into forward contracts at the price level and demand quantities required to fully hedge load or maintain reliability. Then if loads are not hedged during shortage conditions, there will be political pressure to suppress high prices during shortage conditions because the loads are not hedged against the high prices.

If the condition of entry is relatively free the potential for the exercise of market power can also be addressed through forward contracts. Moreover, as discussed in Section III.G below, the same potential for the exercise of market power exists in capacity markets.

Mismatch between Energy and Reserve Pricing and Reliability Requirements

A fourth potential limitation of an energy only resource adequacy mechanism is the potential for mismatch between the incentives provided to resource suppliers in the pricing mechanism and the resource requirements to maintain reliability. If the energy pricing system does not price some element of reliability needs, the required capacity will not be forthcoming. There is therefore a potential in transitioning to an energy only market design for a misunderstanding of the actual operation of the market design or of reliability requirements to result in a gap between the resource supply incented by the pricing system and the resource requirements required for reliability.

Consider for example the second contingency unit commitment for the Boston and Connecticut areas in New England. This is a historical reliability process that attempts to commit sufficient resources on a day-ahead basis so that load in these pockets can be met in real time not only following a single contingency but also following a second contingency. This implies that more capacity will be committed day-ahead than is needed for reliable operation in real-time. If no contingency occurs, there will be no congestion within these load pockets so there will be no premium in the energy market for the capacity located within the pocket. Moreover, any such premium would only accrue to the capacity that is actually dispatched in real-time, but the essence of a second contingency unit commitment is more capacity is committed than is needed to meet load.

The second contingency unit commitment essentially is a requirement for locational reserves. Unlike first contingency reserves, however, this is a soft requirement, no NERC reliability criteria is violated if these reserves are not available; rather the lack of these second contingency reserves is only an issue when the second contingency occurs and it is necessary to reposture the system to sustain the next contingency. The economic value of a second contingency unit commitment could be taken into account in an energy-only market through a sloping demand curve for second contingency reserves within these load pockets.[12] This would provide additional energy/reserve market revenues to resources located within the load pocket when they are used to meet the second contingency requirements. The New England energy market design as implemented in 2003, however, did not have any reserve markets at all, let alone markets for second contingency reserves within these load pockets. Not surprisingly, the energy market failed to provide incentives for the continued operation of existing units within these load pockets needed to meet the second contingency requirement, which are committed uneconomically at minimum load in the energy market, nor for the construction of resources better suited to providing second contingency reserves.

In the ISO-NE case the failure of the energy market design to provide the incentives required for reliability was predictable, but there is also a potential for more subtle differences between the operation of the energy market and reliability needs to give rise to these kinds of risks. These risks are largely transition risks that need to be taken into account in shifting from an effectively functioning capacity market system to an energy market system. Similar, transition risks exist in transitioning from an energy only market design to a capacity market design, there is a potential for the capacity market design to overlook elements of the energy-only market that are necessary to sustain reliability. Moreover, it is unclear whether a capacity market design can identify and sustain the capacity required to maintain reliability in terms of deliverability, the quantity of capacity and the quantity of energy. These potential problems are discussed in Section III. In either case, the risk is initially a reliability risk but probably ultimately a market risk. If the resource adequacy design, either capacity market or energy only, fails to provide the incentives required to sustain the capacity needed for reliability, that capacity will likely be kept in operation by RMR contracts so that reliability is maintained, but the reliability contracts will inflate costs and undermine market incentives. For this reason, there is less risk in transitioning from a capacity market that currently fails to provide the required incentives to an imperfect energy-only market as at worst the market ends up with the same RMR contract problems it already has.

---

12 It would be important to model this reserve as a soft constraint with a sloping demand for these reserves as this constraint cannot always be satisfied.

Conclusion

The potentially important limitations associated with energy only resource adequacy mechanisms are therefore the possibility of unintended changes in the level of reliability during the transition from a capacity-based resource adequacy system to an energy-only system resulting from errors in market design or resulting from misunderstanding of costs levels and the likelihood that regulatory risk will deter needed investment if consumers are not largely hedged against substantial changes in energy prices arising from capacity shortages.

The potential for these kinds of miscalculation and regulatory risks combined with the recognition that there would be little or no price sensitive load in the short run, as well as a reluctance to allow extremely high energy prices, lay behind part of the initial reluctance to rely on energy only pricing to maintain reliability for the initial implementation of LMP markets in New York and PJM.

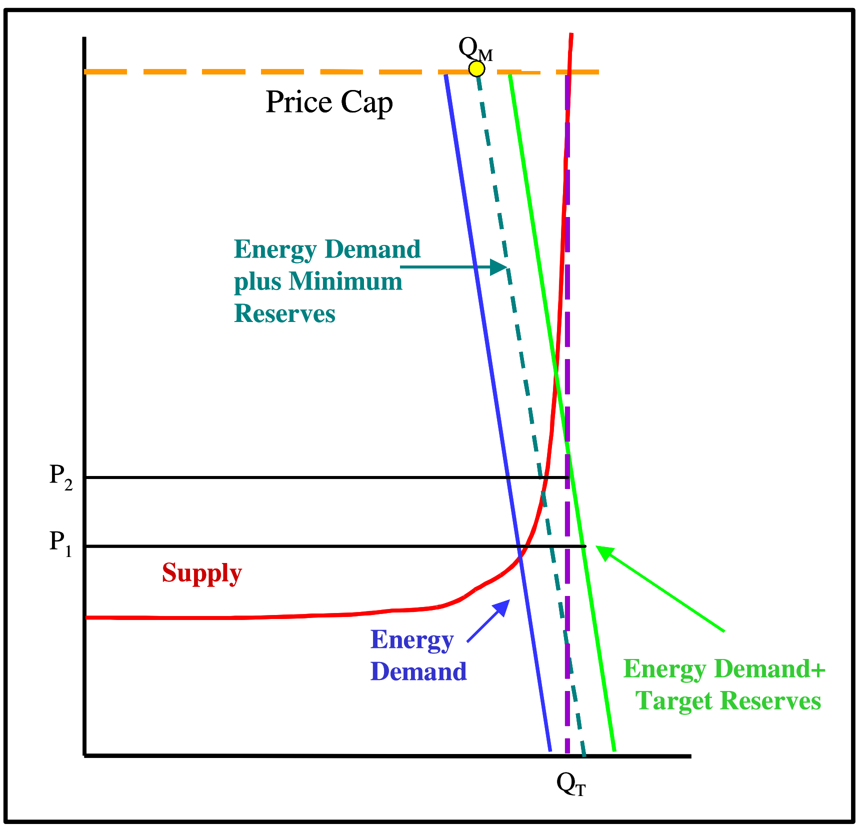
D. Energy-Only Pricing with Price-Responsive Real-Time Load

The potential reliability risks associated with relying on energy only pricing to support resource adequacy in vertically unbundled generation markets discussed above can be reduced if a substantial proportion of the real-time demand for power is price responsive. In an energy-only pricing system with substantial price-responsive real-time load, market clearance and reliability can be ensured by price-responsive real-time customer demand, without the need for administratively determined installed reserve requirements and without undue reliability risk.

- Operating reserve margins would be maintained during high load conditions by price-responsive load reducing consumption in response to high prices.
- There would be no administrative installed reserve requirement or capacity payment. Long-term capacity decisions would be left to market incentives.
- The electricity market would clear while providing reliability, through operating reserve standards, energy pricing and market-determined installed reserve levels.

In a market with a substantial amount of price sensitive load, errors by the system operator or resource supplier in assessing the frequency of shortages as well as flaws in the market design would be of less importance from a reliability standpoint as the errors would result in variations in market prices but would not lead to involuntary load shedding. If there is adequate price-responsive load, even frequent reserve shortages need not lead to involuntary load shedding. The crux of such a system is that rather than shocks such as unexpected weather or unusual levels of generation outages translating into reduced operating reserves and higher load shedding risk, these shocks would result in higher prices that would lead to voluntary load reductions that would maintain operating reserves. Thus, at price P_1 in Figure 9 there is not enough capacity to meet load plus maintain the target level of operating reserves. At price P_2, however, energy demand falls enough to allow the system operator to maintain the target level of operating reserves and avoid entering an emergency state. Moreover, the demand curve portrayed in Figure 9 could shift out considerably, or outages shift the supply curve considerably without the minimum level of energy demand plus minimum reserves (Qm) exceeding the available capacity.

Figure 9
Energy Pricing with Price-Responsive Load



Peak energy consumption and thus the required level of capacity would likely be lower under an energy-only market with price-responsive load than under a capacity market system because consumers could avoid paying for energy whose true cost of production exceeds its value by reducing consumption during high-priced peak periods. The reduced need to support rarely used capacity would be the welfare gain from implementing real-time demand response.

The existence of substantial price-responsive load does not alter the need for effective shortage pricing that results in high prices during reserve shortage conditions. If there is little or no price sensitive load within the price range in which energy and reserve markets are permitted to clear, then errors either by the system operator in assessing the cost of capacity or by suppliers in assessing expected prices could translate under an energy-only pricing system into differences in the frequency of reserve shortage conditions and inefficiently high probabilities of involuntary load shedding.

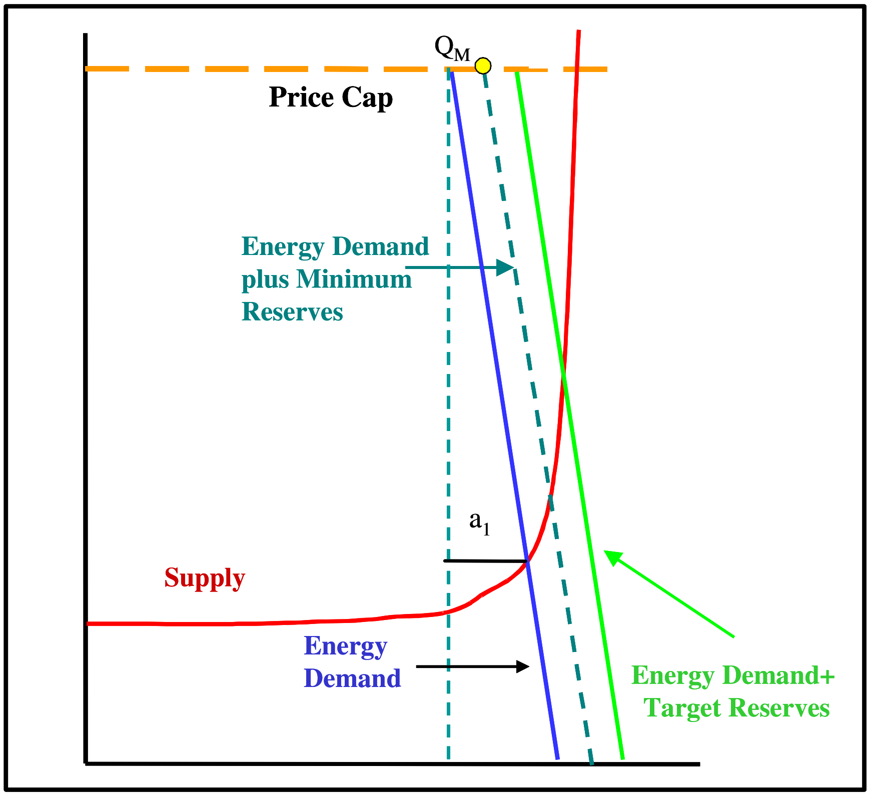
The benefits from developing effective demand response to support an energy only pricing system are therefore twofold: 1) reducing the potential for adverse reliability impacts during the transition from a capacity based resource adequacy mechanism to an energy only market; and 2) avoiding the costs of maintaining capacity to meet the price responsive load. There often seems to be an implicit assumption in discussions of real-time metering and price responsive load that these market features would avoid the need for high prices during shortage conditions or eliminate the regulatory risks associated with energy only pricing. This is not the case.

First, implementation of effective price-responsive demand can reduce the amount of capacity that needs to be kept in operation and thus recover its going forward costs in energy or capacity prices, but it does not change the need for generating capacity to recover its going-forward costs. Second, the magnitude of the benefit from avoiding the costs of maintaining marginal generation depends on the amount of demand reduction that actually occurs under peak conditions. If very little demand response actually occurs, there is very little cost saving. Third, if demand response sets energy prices, implementation of these programs may reduce the number of very high priced reserve shortage hours, but there would be an offsetting increase in the number of hours in which high prices were set by the demand response program. Fourth, implementation of demand response has virtually no impact on the regulatory risk problem, there will still be a potential for years with high and low consumer costs. Each of these observations is discussed in somewhat greater detail below.

First, energy-only pricing systems that utilize price-responsive load to clear the market improve reliability by avoiding the risks arising from mistaken expectations or market design problems under energy-only pricing with price-inelastic demand only if there is in fact sufficient load responding to short-term price signals to enable the market to clear while maintaining reliability levels. Thus, there have to be truly effective demand response programs that can be relied upon to produce real load reductions during high load conditions.

The important factor in assessing the effectiveness of demand response in maintaining reliability is the amount of demand reduction over the price range from the normal price level up to the price cap/shortage price, the range a_1 in Figure 10.

Figure 10
Magnitude of Demand Response

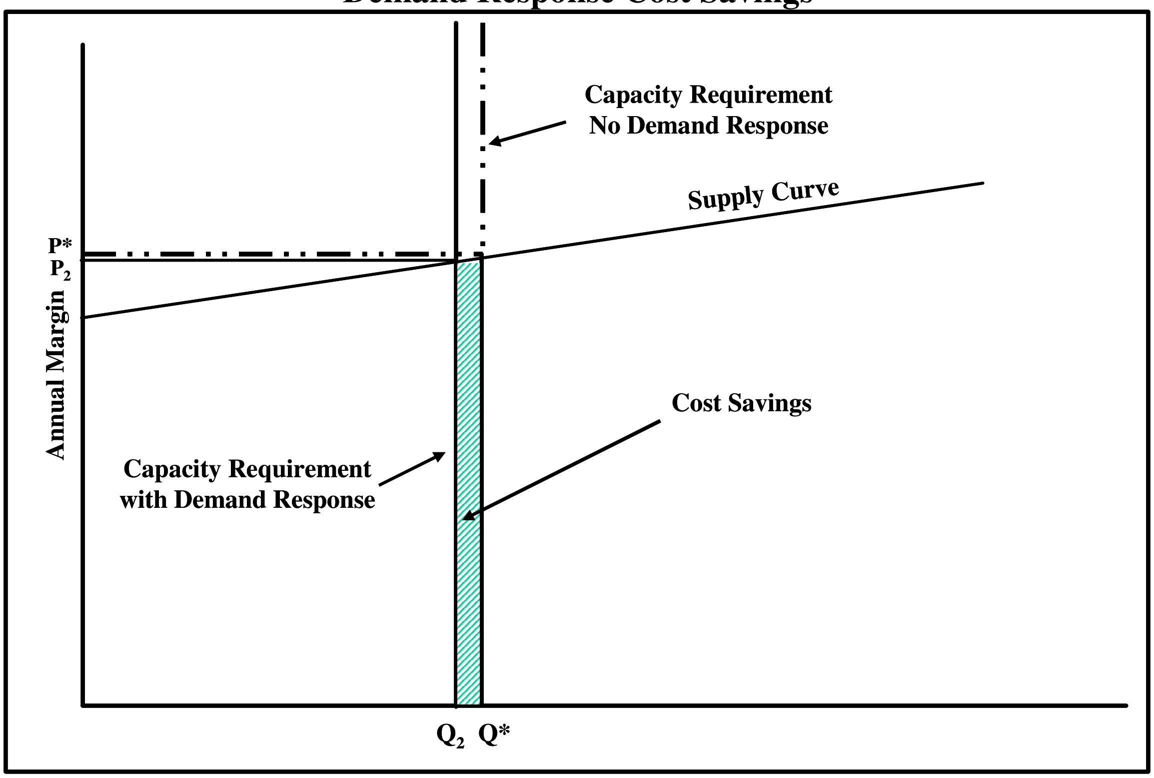


If demand is only somewhat price-responsive, the market equilibrium may be very little different than if demand were completely inelastic. In assessing the practical import of demand response, it is therefore important how price elastic demand is under peak load conditions and how high prices can go under shortage conditions.

This leads to the second observation. Experience to date suggests that only limited demand response is available at low energy prices. State and federal regulators must be willing to allow real-time energy prices to rise well above the incremental cost of the marginal generator in order for demand response to be effective in maintaining reliability. If demand response is called by the system operator during potential shortage conditions but does not set prices, implementation of demand response would have little or no impact on the number of reserve shortage hours in which suppliers recovered their going forward costs, the programs would simply reduce the potential for load shedding. If demand response could set spot prices and were called at prices lower than reserves shortage prices, then implementation of demand response would reduce the number of very high priced reserve shortage hours while increasing the number of high priced non-shortage hours.

Third, the existence of price-responsive load does not change the reality that the marginal generator must be able to recover its going-forward fixed operating costs. If the long-run supply of capacity as a function of expected net annual margin is as portrayed in Figure 11, changes in the amount of capacity needed to meet peak load as a result of demand response programs would not materially change the net margin (P_2 is almost the same as P*) required to sustain the marginal generator. The costs savings from implementation of demand response are the avoided costs of the marginal generation (Q_2-Q*) * P.

Figure 11
Demand Response Cost Savings



Fourth, while the presence of substantial price-responsive load reduces the potential reliability risks arising from misjudgments regarding generation margins and capacity adequacy under an energy-only market, it does not solve the political problem of generator cost recovery entailing high energy prices during a significant number of hours. Moreover, unless demand is very price-elastic or demand peaks very regular, there is a potential for substantial energy price volatility with the recovery of generator fixed costs concentrated in particular years, just as in an energy-only market with little or no demand response.

Thus, even with substantial price-responsive load there will be concerns regarding the potential for the exercise of market power during shortage conditions and concerns regarding the year-to-year variations in consumer costs.

This characterization of capacity and energy markets suggests that the primary advantage of capacity markets is in potentially diminishing the regulatory risk associated with energy only markets in which consumers have not entered into forward contracts and generator recovery of going forward costs is concentrated in particular years. It was observed above that another potential advantage of capacity markets in terms of reducing reliability risks during a transition to energy only markets is really an advantage to a region of staying with its existing resource adequacy mechanism, whether that is a capacity market system or an energy only system, as similar risks exist for a transition from energy-only markets to capacity markets. Moreover, it has been suggested above that there are practical problems in actually maintaining reliability under capacity market systems. Importantly, under either an energy-only market, a capacity market system or a system in which load is met with the regulated generation of the vertically integrated utility, consumers in the end must pay the going-forward costs of existing generation as well as a return of and on investment for new generation. Offset against the potential advantage of capacity market systems in reducing regulatory risk, a capacity market system can have a wide variety of practical implementation issues that may be better managed or avoided entirely under an energy-only pricing system. These issues in implementing capacity market systems are discussed in Section III.

III. CAPACITY MARKET DESIGN ISSUES

Most of the turmoil in Eastern capacity markets since 1998 has revolved around basically the same set of issues in each of these markets. The description of the Eastern capacity markets below is organized around these core issues, explaining how these common issues have been addressed in each market and describing and discussing how some of the proposals for changes in the design of these markets would address these problem areas.

A. Defining Capacity Requirements

A necessary component of any capacity market is a process for determining the amount of capacity that LSEs will be obligated to purchase. This determination has several elements.

1. Reserve Margin

The New York State Reliability Council determines on an annual basis the installed reserve margin required for the NYISO to satisfy NPCC Resource Adequacy criteria of a one day in ten years probability of shedding firm load due to inadequate resources. This is a probabilistic Monte Carlo analysis that takes account of scheduled and forced outages and deratings of capacity resources, availability of imports from neighboring control areas, and capacity or load relief available through operating procedures.[13]

The NYISO then determines the minimum installed capacity requirement for the NY control area as the product of the forecasted control area peak load and 1 plus the reserve margin. This capacity requirement is stated in terms of the total rated capacity of the capacity resources. The NYISO translates the minimum installed capacity requirement into what is effectively an expected amount of capacity (the minimum unforced capacity requirement, UCAP), discussed more fully in Section C) by multiplying the installed capacity requirement times 1minus the average outage rate (EFORd) value for the six most recent 12-month rolling average outage rate of all New York resources in the New York control areas.[14] The NYISO also determines the locational minimum installed capacity requirements for Long Island and New York City.[15]

PJM's process for determining its current installed capacity requirements is similar to New York's. Electric distribution companies submit their load forecasts for the planning period to the PJM. PJM uses distribution company forecasts as well as historic peak load information to determine zonal peak load forecasts for the planning period.

PJM and PJM's Reliability Assurance Agreement Reliability Committee determines the PJM reserve margin. This reserve margin is determined through a probabilistic analysis that accounts for both load forecast uncertainty, maintenance outage requirements and uncertain generator forced outages based on historic outage rates (EFORd). The current expected loss of load probability target is one day in ten years.[16] The historic EFORd ratio is then used to convert the nominal capacity requirement into an expected capacity requirement (UCAP).

ISO-NE has used a similar one day in ten year loss of load criteria based on similar analysis of scheduled and forced outages, assistance from external control areas and capacity or load relief from operating procedures to determine the NEPOOL reserve

---

13 NYISO Installed Capacity Manual, p. 2-3; New York State Reliability Council LLC, Policy No. 5-0, Procedure for Establishing New York Control Area Installed Capacity Requirements, August 11, 2003; New York State Reliability Council, LLC, Installed Capacity Subcommittee, New York Control Area Installed Capacity Requirements For the Period May 2005 through April 2006, December 10, 2004, Appendix A (hereafter NYSRC 2004).

14 NYISO Installed Capacity Manual, p. 2-3.

15 NYISO Installed Capacity Manual, p. 2-3.

16 PJM Manual 20, pp. 17-22. PJM Reliability Assurance Agreement, May 17, 2004, Schedules 4, 4.1.

margin.[17] As in PJM and NYISO, a historic outage (EFORd) ratio is then used to convert the nominal capacity requirement into an expected capacity or UCAP requirement.

A common feature of this process for determining the capacity requirement is that it effectively starts with a known set of capacity resources with known expected outage rates and from this determines the amount of capacity needed to meet the reliability standard given the characteristics of these resources.[18] There is a slight logical disconnect in this process of determining the capacity requirement because the process of determining the capacity requirement must start with an assumption regarding which resources will ultimately be selected to provide capacity in order to assign outage rates to the resources used in determining the capacity requirement. This disconnect does not appear to have yet had material consequences under the current capacity market designs, perhaps because there have not been wide differences in outage rates among most of the resources and the set of resources used to provide capacity has not changed much from year to year. Nevertheless, if there were differences in the outage rates of the resources that would provide the marginal capacity and it were not known which resources would actually be selected to provide capacity, there would be a potential mismatch between the expected availability of the resources used to determine the capacity requirement and the expected availability of the resources actually selected to meet the capacity requirement.

The significance of this potential disconnect will more important as the mix of capacity resources changes over time to include resources with more diverse ratios of expected to nominal capacity. As long as the set of resources that will actually be providing capacity is largely known at the time the capacity requirement is determined, resources with diverse outage rates can be taken into account in the process of determining the capacity requirement. Changes in market design that make it less clear which resources will actually be selected to meet the capacity requirement, such as changes that push the capacity procurement forward in time to an earlier stage in the development cycle will make this disconnect more important.[19]

Another respect in which this distinction between nominal and expected is important is in determining the amount of capacity that is paid for by LSEs. The basic principle is simply that consumers should be paying the same price (net of performance adjustments) to resources providing the same amount of expected capacity, rather than to resources providing the same amount of nominal capacity.

---

17 ISO New England, Manual for Installed Capacity, p. 1-4 and 1-5.

18 Under a UCAP system, this gross capacity is then converted into a requirement stated in terms of expected capacity that is purchased in the auction.

19 This issue is discussed below in the context of resource availability incentives and the ISO-NE FCM proposal.

2. *Allocating Capacity Requirements to LSEs*

Once the aggregate capacity requirement for each control area has been determined for the forthcoming year, it is allocated to LSEs.[20] Under the original capacity systems implemented by Northeast ISOs, LSEs purchased capacity to meet their assigned requirement or paid a deficiency charge for the shortfall in their purchases. This deficiency payment was loosely related to the cost of a gas turbine. The current PJM deficiency charge is $160/MW day, $58,400 MW year, divided by one minus the average EFORd.[21]

New England

New England's procedures for allocating capacity responsibility to LSEs are the simplest. Capacity requirements are allocated to each load in proportion to that load's share of the current year's system peak load. System peak load is measured as load during the single highest load hour for the ISO-NE control area. Each LSE's capacity requirement is then the sum of the capacity requirements allocated to the loads it serves.[22]

PJM

PJM uses a slightly different procedure to measure system peak load and reflects anticipated load growth in its allocation mechanism. Capacity requirements are allocated among zones (i.e., areas served by a single utility), in proportion to each zone's forecasted share of the forthcoming year's system peak load. System peak load is measured as the average of loads during the five highest load hours for the PJM control area. Each zone's share of capacity requirements is then allocated to loads in that zone in proportion to each load's current-year contribution to system peak load.[23]

New York

New York also reflects anticipated load growth in its allocation, but it does not allocate requirements based on shares of system peak load. Instead capacity requirements are allocated among transmission districts, which are similar to PJM's zones, in proportion to the forecast for each transmission district's individual peak load for the forthcoming year. The transmission district peak is measured as load during the single hour in which load in the transmission district is highest. Each transmission district's share of capacity requirements is then allocated to loads in that transmission district, in proportion to each

---

[20] To be precise, it is a UCAP requirement that is assigned to LSEs. UCAP is explained in Section C.

[21] PJM Reliability Assurance Agreement, May 17, 2004, Schedule 11.

[22] ISO-NE ICAP Manual, Section 2.1.

[23] PJM Manual 20, Reserve Requirements, Section 2, April 30, 2004.

load's forecast contribution to that transmission district's peak load, based on its actual contribution to peak load in the prior year.[24]

3. Demand Response

Capacity systems need to have mechanisms to account for and provide incentives for demand response programs, since capacity market systems depress energy prices during reserve shortages reducing or eliminating the incentive for demand to reduce consumption during peak load conditions.

PJM has taken a somewhat different approach to demand response than ISO-NE and the NYISO. PJM's peak load forecast takes account of active load management capability, so qualifying demand response in effect avoids capacity charges.[25] The degree of capacity credit provided to active load (ALM) management programs is determined in the probabilistic reliability analysis and can vary between 0 and 1.[26] A distinguishing feature of this system is that it is up to the LSE serving the load to provide demand response in order to reduce its capacity charges.

In New York demand response resources can qualify as special case resources which count as capacity, i.e., they count as capacity that can be sold to LSEs to satisfy an LSE's capacity requirement.[27] Like generators providing capacity, Special Case Resources must be capable of reducing load for a minimum four hour block. Analogous to the obligation of generators to bid in the day-ahead market, Special Case Resources are notified day-ahead that they may be needed.[28] During the operating day they will have two hours' notice to reduce load. At present Special Case Resources are also paid in the energy market for their demand reductions.[29]

Unlike PJM, in New York the entity providing demand response need not be the entity serving the end use customers load. This is an intentional feature motivated by a perception[30] that the distribution companies that are the LSEs serving most load are not

---

24 NYISO ICAP Manual, Sections 3.3 and 3.4. There are also detailed rules governing customer switching (Sections 3.2 to 3.10).

25 PJM Manual 20, p. 11-14. PJM Reliability Assurance Agreement, May 17, 2004, Schedule 5.2.

26 PJM Manual 20, p. 25-26.

27 NYISO ICAP Manual, Section 4.12.

28 The NYISO is generally obligated not to use this day-ahead notification "indiscriminately" but "only when the Day-Ahead Market indicates serious shortages of supply for the next day." NYISO ICAP Manual, p. 4-31.

29 NYISO ICAP Manual, Section 4.12.8.

30 This perception may or may not be correct.

interested in nor well suited to providing demand response and would fail to develop these opportunities.

ISO-NE's approach to demand response is similar to New York. Demand resources in the Real-Time Demand Response or Real-Time Profiled Demand Response programs can qualify as capacity resources and sell capacity in the capacity market.[31] ISO-NE demand response must be able to interrupt upon two hours notice, without regard to day-ahead notification.[32]

Both the NYISO and ISO-NE demand response programs have a variety of additional rules for measuring load and load reductions that are beyond the scope of this discussion. Many of these rules relate, however, to two features of demand-response mechanisms within capacity market resource adequacy mechanisms that are relevant in assessing capacity market systems. First, while demand response in an energy market system is incented by the cost savings from avoiding the need to pay high energy prices during shortage conditions, capacity market demand response programs must pay demand response to reduce consumption. This requires that demand response be measured relative to some benchmark level of consumption in capacity market systems in order to determine the amount of reduction to be paid for. Since the consumption of a demand response load in practice varies with many factors, there is a likelihood under capacity market systems both of excessive payments for demand response at times when consumption is low for other reasons, and inadequate payments for demand response at times when consumption is high for other reasons. Thus, if a load has a benchmark demand of 10 MW and offers 3 MW of demand response, the load has no incentive to reduce its consumption on a day when its normal consumption would be 7 MW. Conversely, on a day when its normal consumption would be 12 MW, it would need to reduce its load by 2 MW, just to get down to its benchmark, so it might be too expensive to even try to provide demand response.[33] Capacity market systems therefore need a variety of rules to measure the benchmark level of demand, but there is no ideal solution to this benchmarking with the capacity market design.

Second, demand response programs in capacity market systems potentially impose an undefined obligation on the load offering demand response. If the demand response resource must reduce consumption every time it is called, the cost of providing the demand response depends on how often the resource will be called. The cost of

---

31 ISO New England Load Response Program Manual, Section 7.

32 ISO-New England has two categories of real-time load response, 30 minute response and 2 hour response. The 30 minute load response is paid a higher price for its energy reductions. See ISO New England Load Response Program Manual, Section 2.2.

33 The problems associated with a benchmark that is too low could be avoided by setting the benchmark for demand response at the level of the maximum potential consumption, but this would often leave little incentive for demand response because the normal consumption level would be less than the benchmark less the specified demand response.

providing demand response for 10-20 hours a year may be much different than the cost of providing demand response 200 hours a year. This was seen in California where many of the traditional demand response resources did not want to participate in the program during 2001, after having been called a very large number of times in 2000 and expecting even worse for 2001. This problem is exacerbated by the fact that the frequency with which demand response resources will be called depends not just on weather but on the performance of other resources in the capacity market system. Analyzing the potential performance obligation under a capacity market system would therefore require that the demand response resource assess the potential for substantial outages of other resources and the accuracy of the demand forecast, not merely its own ability to reduce consumption.

B. Deliverability

1. Overview

Capacity deliverability tests are a central issue in implementing capacity market systems in decentralized electricity markets, particularly with respect to the ability of new generators to participate in the capacity market. PJM, NEPOOL and the NYISO rely on locational energy pricing for congestion management. This has enabled all three ISOs to adopt a "minimum interconnect" standard for generators selling energy into the market. A new generator satisfies the "minimum interconnect" standard if it is able to deliver its power to the transmission grid without adversely affecting reliability and its interconnection (at zero energy dispatch) does not reduce transfer capability.

LMP pricing in energy markets provides new generators with incentives to site themselves efficiently, without restricting competition. Congestion impacts are reflected in the locational energy prices and thus in the revenues of both incumbents and entrants. Generators that locate at places where they often cannot be dispatched because of transmission constraints will earn low energy margins under LMP pricing. The prospect of low margins due to congestion thus serves to incent new generation to locate where capacity is needed and energy prices are higher.

Generators receive capacity payments, however, whether they operate or not, so there is no locational price signal in the capacity market absent some form of deliverability requirement.[34] Absent any form of deliverability requirement there is a potential for capacity to be developed in locations at which it is cheap to construct, even if, because of transmission constraints, the capacity adds little to the amount of power that can be used to meet load under stressed system conditions. The more important the

---

[34] As discussed below, deliverability requirements can take many forms, ranging from the locational capacity requirements of the NYISO to the CETO/CETL tests of PJM.

capacity payment is as a source of generator revenue, the greater the potential incentive problem. Thus, if a substantial proportion of the net margin of the marginal generator is derived from the energy market, it is less important to impose deliverability requirements in the capacity market as capacity that is not dispatchable to meet load under stressed system conditions will likely be uneconomic regardless of whether a deliverability test is applied for capacity market purposes. The larger the proportion of revenues of the marginal generator that are derived from the capacity market, however, the greater the potential, absent a capacity market deliverability requirement, for construction and continued operation of generation that is not cost effective in terms of its contribution to regional reliability.

All three Northeast ISO's have struggled with how to apply some form of deliverability test to sellers in the capacity market and have taken different approaches to resolving this problem. Such a test should satisfy at least three objectives.

- No barriers to entry: The deliverability test should preserve the condition for efficient entry to be profitable if the entrant's full generating costs are less than the avoidable generating costs of the incumbent.

- Permit long-term capacity contracts: The deliverability test should permit long-term bilateral contracts for capacity. This requires that capacity sellers be able to hedge themselves against the impact of future entry of new capacity on the deliverability of their capacity resource.

- Reflect reliability criteria: The deliverability test needs to ensure that resources eligible for capacity payments make an appropriate contribution to reliability under stressed system conditions.

2. PJM

The PJM deliverability requirement for capacity market resources tests whether the aggregate of capacity resources can be utilized to reliably deliver energy to aggregate control area load. This deliverability requirement has two components based on probabilistic load and outage analyses. First, the ability of an electrical area to export energy to the remainder of the control area is tested to ensure that the capacity resource is not bottled. This test ensures that each electrical area is able to export any surplus capacity at peak load (i.e., under stressed system conditions). The failure of a new resource to pass this tests implies that it is bottled and that additional transmission must be built for this capacity to be deliverable to PJM load outside the subregion.[35]

[35] Attachment E: PJM Deliverability Testing Methods.

This deliverability requirement has two important features. First, all generation subject to the test fails or all generation passes. Second, existing capacity market resources are grandfathered so the failure of capacity resources to collectively satisfy the deliverability test does not affect the ability of incumbents to supply capacity; it only excludes competition from entrants. Suppose, for example, that PJM determined that 1,000 MW of capacity could be exported from a particular generation pocket and 1,000 MW of capacity existed that had previously been approved as capacity market resources. A new entrant would not be approved as a capacity market resource unless it expanded the transmission system to satisfy the deliverability requirement and thus would be unable to undercut the incumbent capacity suppliers, even if the entrants full costs were considerably lower than the market price of capacity demanded by the incumbent suppliers. This grandfathering of incumbents allows capacity resources to enter into multi-year capacity contracts but violates the efficient entry condition described above.

The second reliability test assesses whether energy will be deliverable from the aggregate of PJM resources to the load in the portions of a PJM subregion experiencing a localized capacity deficiency. The second test is based on the Capacity Emergency Transfer Objective (CETO) and Capacity Emergency Transfer Limit (CETL) tests, applied to electrical subareas within PJM. The CETO measures the amount of energy that the subarea must be able to import in order to remain within MAAC reliability criteria. The test is passed if the actual emergency transfer limit (CETL) exceeds the CETO for the area. This deliverability test was historically applied to the service territories of the member Investor Owned Utilities.[36] More recently, PJM has begun to apply the test to other electric subareas within these service territories.[37] While the application of the CETO/CETL test is reasonably clear, the consequences of failure are not because it is the region, not a resource, that fails.

PJM's "Deliverability Testing Methods" state that "Failure of deliverability tests brings at least two different possible consequences. When evaluating a new resource, if the addition of the resource will cause a deliverability deficiency then the resource cannot be granted full capacity credit until system upgrades are completed to correct the deficiency. If the deliverability of PJM degrades, for any number of reasons, failure of deliverability tests may result in a sub-area being unable to receive full capacity credit for remote capacity resources delivered to that subarea."

In the circumstance in which the exit of a generator or load growth cause a load pocket to fail the CETO/CETL test, the PJM deliverability test does not appear to address

[36] This made sense from the perspective of the historical role of the required reserve margin as reliability problems arising from transmission constraints internal to the service territory of a single utility would be internalized by the utility having the obligation to serve that load and would not impact other pool members.

[37] See PJM, PJM Reserve Requirements, Manual M-20, April 30,2004, pp.23-24. PJM Attachment E: PJM Deliverability Testing Methods.

the question of which resources can qualify as capacity resources for LSEs within the region and which LSEs must bear the financial burden of contracting for potentially high cost generation within the region to satisfy the CETO/CETL test. Does failure to contract for high cost generation in a load pocket trigger transmission expansion?

The PJM capacity market deliverability requirement is workable for generation potentially located in generation pockets, as generation sited at such a location would be required to pay for transmission upgrades to expand deliverability if the new capacity otherwise would not make a sufficient contribution to PJM reliability. As noted above, however, the grandfathering of existing capacity under this system potentially deters efficient entry and keeps high cost generation in operation. This capacity market deliverability requirement works less well for incenting generation to locate within high cost load pockets and can break down if such load pockets exist. If a capacity shortage develops within any load pocket within PJM, no new generation might be able to meet the capacity market delivery test unless the shortage within the load pocket were eliminated. The cost of siting generation within the load pocket, or of a transmission expansion to deliver capacity into the load pocket might greatly exceed the market price of capacity elsewhere in PJM, so which LSE must pay for this marginal capacity? In such a circumstance, no market participant would be willing to incur the cost of siting generation or building transmission to relieve the shortage in the load pocket, but until such capacity was added, the region could not pass the CETO/CETL test.

Suppose, for example, that the cost of new capacity required to satisfy the deliverability requirement within a particular load pocket were $150,000/MW, but the going-forward price of capacity located outside the load pocket was only $5,000/MW. Under the PJM procedures, no new resource outside the load pocket could satisfy the deliverability test so there would be no ability to undercut capacity prices demanded by grandfathered incumbents up to $150,000/MW. Moreover, the LSE actually serving load inside the load pocket could have entered into long-term contracts for capacity with grandfathered resources located outside the load pocket and escape the consequences of the high capacity market prices brought on by the exit of capacity located within the load pocket.

The combined effect of these features could be to exclude all new resources from the capacity market and to prevent new resources from undercutting incumbent offers unless the entrants fund transmission investments sufficient to reliably deliver energy to aggregate control area load throughout PJM, but the cost of such investments could exceed the price of capacity.

In practice, these patterns have not appeared in the PJM capacity market. For reasons that are not always apparent, prices in the PJM capacity market have been set at very low levels by generation located outside the Eastern PJM load pocket while a large number of units in Eastern PJM are apparently unable to remain in operation at these capacity market prices and have been seeking to exit the market, yet their continued operation would be necessary in order for generation in PJM to satisfy the aggregate deliverability test. These problems in the capacity market provided part of the impetus for the development of a new capacity market design for PJM, called the Reliability Pricing Model.

3. New York

The New York capacity market system was developed with the Manhattan and Long Island load pockets in mind and with the intent of placing all resource providers on a level playing field. Rather than imposing a control area-wide capacity market deliverability requirement, New York has attempted to ensure deliverability by establishing locational capacity market requirements. LSEs serving load in New York City are required to procure at least 68 percent of their capacity requirements (or 80 percent of peak load) from NYC resources. LSEs serving load in Long Island are required to procure at least 84 percent of their capacity requirements (or 99 percent of peak load) from Long Island resources.[38] This system allows capacity prices in an individual load pocket to rise to the level required to warrant new investment or to keep existing capacity in operation. The advantage of this locational system relative to a PJM-type deliverability requirement is that the price of capacity within a particular load pocket (Manhattan) can be very high at the same time that capacity prices are much lower elsewhere (upstate New York) and incumbents and entrants have equal access to the transmission system.

New York has also developed the concept of unforced capacity deliverability rights (UDRs) for new transmission projects that enable power to be delivered into New York City or Long Island from capacity located elsewhere. The construction of additional transmission into Long Island or New York City will not change the locational capacity requirement. Instead, the transmission project would be awarded UDRs reflecting the ability of the transmission assets to deliver additional power into the capacity market region. A UDR combined with an upstate capacity resource would then count as in city or on island capacity.[39]

The New York locational capacity market system has several potential limitations. Perhaps the most important limitation is that the New York City locational capacity market tended to clear at the price cap set in the Con-Ed divestiture contracts, so the price cap for divested generation acted like an administratively set capacity price.[40] A second limitation is that there could be additional load pockets within the New York

---

[38] NYISO Installed Capacity Manual, Attachment B, April 26, 2004. These percentages are not fixed and can change for each capability period.

[39] See NYISO Services Tariff, Section 5.11.4, Sheets 127-127A; NYISO ICAP Manual, Section 4.14, p. 4-34; and Internal NYISO DC Controllable Line Scheduling, Concept of Operations, May 4, 2004, p. 8.

[40] This situation has changed with the implementation of the demand curve discussed in Section C.

transmission system that are not represented in the capacity market system. For example, there is no east of Central East capacity requirement, yet the Central East transmission constraint can prevent Western generation from being used to meet Eastern load during shortage conditions. There is also a possibility of reliability problems within Long Island or within In City load pockets that would not be reflected in the On Island or In City requirements.[41]

4. NEPOOL

ISO-NE proposed and filed at FERC in early 2004 a locational capacity market system that is similar to the system in place in New York. Under the original proposal, NEPOOL would have four locational capacity market regions, Connecticut, Northeast Massachusetts and Boston, Maine and rest of New England.

It is noteworthy that generation ownership in these proposed regional capacity markets is much more concentrated than in New England as a whole. While the HHI for generating capacity in New England is around 750 according to ISO-NE, the HHI in Connecticut was 2300 and over 5300 in Boston.[42] This proposal was subsequently modified in an August 2004 proposal that included Southwest Connecticut as a separate zone.[43] The FCM proposal as of Spring 2006 retains a zonal structure in principle, although no zones are specified in the Settlement Agreement. Instead, the Settlement Agreement describes criteria to be used in determining when a region will be modeled as a separate zone in the capacity auction.[44] This test is essentially whether existing capacity internal to the hypothetical zone exceeds the local capacity requirement. If the load forecast is growing over time, any region that becomes constrained in the auction will remain a separate zone, as the existing capacity would not be sufficient to meet the load growth. If economic conditions cause a decrease in the load forecast, however, this criteria could cause an existing capacity zone to not be modeled in the auction, yet a substantial capacity price premium might be required to keep the target level of capacity available inside the zone. If this occurred, the resulting capacity shortfall would trigger

---

41 The NYISO transmission owners have recently made a filing at FERC calling for the addition of a deliverability requirement for capacity resources; see Request for Clarification, Request for Extension of Compliance Deadline and Request for Waiver of Notice of Requirements of the New York Transmission Owners, Docket Nos. ER04-449-003, ER04-449-007 and ER04-449-008, May 4, 2006.

42 ISO-NE March 2004, p. 48.

43 Prepared Direct Testimony of Steven Stoft, FERC Docket No. ER03-563-030, August 31, 2004 (hereafter Stoft 2004).

44 Settlement Agreement, Item 11, Section III.A.

the identification of the region as a capacity zone in the subsequent reconfiguration auction.[45]

The various versions of the NEPOOL locational capacity market system included Capacity Transfer Rights (CTRs) across each capacity market interface. CTRs would be a financial instrument that hedged inter-market capacity price differences. They would entitle the holder to the difference in the price of capacity in the capacity spot auction between the two capacity regions specified by the CTR. Entities holding CTRs would therefore receive a payment reflecting locational capacity market price differentials. CTRs would therefore be financial instruments like FTRs, but they would hedge inter-regional differences in capacity prices when constraints on inter-regional capacity transfer were binding in the capacity market auction.

In the initial LICAP filing, CTRs were to be allocated monthly to loads inside Boston and Connecticut reflecting the existing transfer capability. CTRs out of Maine were to be allocated proportionately to resources located in Maine.[46] In addition, there would be a special allocation of CTRs to municipal utilities reflecting their historic entitlement to use of the transmission system.[47] Thus, loads inside the Boston and Connecticut pockets would buy a proportion of their capacity at the rest of New England capacity price and generators in Maine, an export constrained region, would sell a portion of their capacity at the rest of New England price.[48] These special CRT allocations were eliminated in the Fall 2004 LICAP filing.[49] The allocation rules for CTRs changed somewhat as the LICAP proposal evolved into FCM but the basic structure remained the same.[50]

In the initial LICAP filing, the capacity prices determined by the capacity market auction were to be capped in the import constrained regions with a five-year phase-in, with the cap rising by $1,000/MW month per year.[51] Offsetting this cap would be a $5,340/MW month transition payment to generation in the Boston and Connecticut load zones with a capacity factor of less than 15 percent in 2003. These transition payments

---

45 This would also happen in newly binding zones; if insufficient capacity cleared in the three-year out auction, the zone would be enforced in the subsequent reconfiguration auction.

46 ISO-NE, March 2004, p. 39.

47 ISO-NE March 2004, pp. 39-40.

48 This rule in effect transfers part of the value of the transfer capability between Maine and the rest of New England to Maine generators, despite the fact that Maine generators do not pay the embedded cost of this transfer capability.

49 Mark Karl, Prepared Direct Testimony, Docket No. ER03-563-030, August 31, 2004, pp. 22-23.

50 Settlement Agreement, Item 11, Section III.A.6.

51 ISO-NE March 2004, pp. 6, 27-29

would be borne by network load in the import constrained subregion.[52] The details of the transition payments have evolved as the original LICAP proposal has evolved into the FCM proposal but a transition period with transition payments continues to be part of the proposal.[53]

The overall capacity requirement for NEPOOL would be determined and then allocated to the capacity regions based on peak loads during the prior year. The ISO would then determine the minimum level of capacity required within each region by removing capacity within a region until the reliability criterion was violated.[54] The March 2004 ISO-NE locational capacity proposal explicitly provided for cascading of locational capacity prices.[55] This is important for ensuring rational prices when excess capacity is offered within a load pocket.

5. Comparisons and Extensions

An important advantage of a locational capacity market system relative to a PJM-type deliverability system is that it permits a cost premium to be reflected in the capacity payment to generation located within high cost load pockets, making it economic for such capacity to remain in operation despite lower capacity prices in other regions. This feature of a capacity market system also ensures that a control area wide capacity shortfall does not result whenever it becomes uneconomic to build new capacity within one or more load pockets at the regional capacity price.

One reality of a locational capacity market system is that a capacity requirement for highly concentrated load pockets combined with an administratively determined capacity price cap essentially amount in the short run to an administratively determined capacity payment, with very little role for markets. Thus the Con Ed locational capacity payment generally cleared at the price cap set in the Con Ed divestiture contracts. This kind of outcome is even more likely in smaller load pockets with even fewer competing suppliers. In the long run in which loads can contract with entrants for new capacity, there is much more potential for competitively determined capacity prices, even in concentrated load pockets, but this requires that capacity market buyers have long-term load-serving obligations that permit them to enter into long-term capacity contracts, a

52 ISO-NE March 2004, pp. 6, 29-32 Units receiving the transition payments will be subject to a variety of restrictions, in particular, tighter mitigation of offer prices than other units After the phase in, the capacity market price cap will be removed in the constrained regions and all units will be subject to this tighter mitigation. ISO-NE March 2004, p. 49.

53 Settlement Agreement, Item 11, Section VIII.

54 This requirement would then be translated into UCAP under the March 2004 filing and the locational UCAP requirement would be subtracted from the regional UCAP requirement to determine the amount of UCAP that could be imported into each capacity market region.

55 ISO-NE March 2004, pp. 35-39

topic which is discussed further in Section C below. The NYISO capacity market demand curve discussed in Section C addresses this situation to a degree by allowing locational capacity prices to vary in a range with changes in capacity supply.

There are a variety of potential variations of the NYISO/ISO-NE locational capacity market systems to address the problem of intra-zonal generation pockets. One approach would be to add an intra-zonal deliverability requirement to the locational system. Thus, generation located on Long Island would only qualify as capacity if it satisfied an intra-Long Island deliverability requirement. Such an approach could work if the goal is to ensure that developers do not site a disproportionate amount of new generation within a generation pocket and could incent developers to spread out new generation. Such a system could, in principle, however, deter efficient entry. Moreover, such an approach would not work as well, and have limitations similar to the PJM system, if reliability required incentives for new generation to locate within specific load pockets within a zone.

Second, by combining a locational capacity market system with shortage pricing within load pockets, one could use energy market revenues to improve the incentives provided by a capacity market system. It was noted above that the locational incentives provided by a capacity market system become more important as the proportion of revenues that marginal generator receives from the capacity market system rises. If the marginal generator will earn substantial revenues from the energy and reserve markets if it locates within constrained load pockets within a zone, this would tend to limit the need to reflect these incentives in the capacity market system.

Even if the marginal capacity resource does not recover a substantial portion of its going forward costs in energy and reserve markets, these markets can be important to the workability of a zonal capacity market design by enabling resources located in load pockets within the zones to earn additional revenues that ensure that needed resources are infra-marginal. This requires, however, that the relevant energy and reserve markets exist and that they operate competitively, with rents that reflect shortage conditions rather than the exercise of market power.

A third approach to addressing intra-zonal constraints would be to use availability performance incentives in the capacity market to measure deliverability. This alternative is discussed in section E in conjunction with the availability incentives proposed as part of SO-NE's LICAP and FCM models.

A particular problem in using capacity and energy markets to sustain the level of capacity needed to maintain conventional reliability standards arises in load pockets in which the unit commitment is based on a second contingency, but there is no reserve market reflecting this requirement. Within these load pockets, such as Boston and Connecticut in New England, the ISO seeks to commit enough resources day-ahead so that load in the pocket can be met even if two contingencies (two independent transmission or generation outages) were to occur. Since both day-ahead schedules and the real-time dispatch are secured only against the first contingency, there will generally be no energy market premium within these load pockets. Moreover, since the extra resources within the load pocket are needed for local reliability rather than transmission system security they are not hard NERC-type reserve constraints but are typically treated as soft constraints in real-time operation, so no value is assigned to real-time reserves within these load pockets. Absent the development of some kind of reserve market to price these second contingency resources, it is predictable that such second contingency commitment areas will need to be modeled as zones for capacity market purposes.

C. Forward Contracts and Retail Access

1. Overview

Under power pool operation, both PJM and New York had installed capacity requirements imposed on each of the vertically integrated transmission owner LSEs belonging to the pool. Each transmission owner had an obligation to serve load in its service territory so they were able to anticipate their installed capacity requirements and to add capacity to satisfy the annual requirements when needed to meet load growth. The potential for the exercise of market power by generation-long utilities was constrained by the ability of pool members to anticipate future installed capacity requirements and to add quick-start capacity or return mothballed units to operation to meet installed capacity requirements within the planning horizon.

This system is fundamentally altered under the retail access programs implemented in many Northeastern states as LSEs under retail access systems do not know what load they will be serving on any future date and often have only short-term contracts with their retail customers. Moreover, there is a potential for individual LSEs to satisfy their individual capacity market requirements by releasing customers if capacity prices are high, dumping the capacity requirement upon the provider of last resort. In addition, in an environment with retail choice, capacity markets must incorporate mechanisms to accommodate load switching between LSEs without undermining the reliability role of the capacity requirement. Critically, compatibility of capacity market systems with retail access requires a mechanism for settling day-to-day imbalances in capacity market obligations as load shifts between LSEs.

The need for a daily balancing system in the capacity market is a potential problem as suppliers do not take capacity out of service or put it in service on a day-by-day basis so prices in these daily capacity imbalance "markets" may not reflect market forces. Suppose, for example, that LSE A contacts for 250 MW of capacity with four different suppliers, paying $36,600/MW year for the capacity, an average of $100/MW day. If LSE A loses 50MW of load to LSE B on June 1, what is the market value of this capacity? On the supply side, what would the marginal capacity market seller pay to buy back the capacity market obligation from LSE A so that the supplier could close its plant and avoid its going-forward costs for the rest of June? The likely answer is not much. The $36,600 in going-forward costs on an annual basis is not incurred on a daily basis, $100/MW day, but is incurred in chunks and once incurred is sunk. If the seller has already taken his unit off-line and incurred maintenance costs, how much will the seller save by mothballing the unit during June? By June, the marginal capacity market seller has probably already entered into a variety of contracts that it cannot cancel on a monthly basis. Even labor costs may be sunk to an extent, for if the generator lays off its employees in June, it may not be able to rehire them in July, so labor costs also may not be variable on a daily or even monthly basis.

Conversely, if LSE A refuses to sell the capacity credits it no longer needs to LSE B, what is the supply price at which additional capacity resources would enter the market to supply these credits? If the next highest cost capacity supplier at the beginning of the year offered capacity credits at $ 36,966/MW year ($101/MW day), this supply would very likely not be offered for the month of June for $3,030/MW. That marginal supplier might not be even able to bring its resource out of mothballs soon enough to be available during June and, even if it did, it might need to incur nearly the same $36,966 to be available in June that it would have incurred in order to be available for the year.

The only significant supply-side factor affecting daily capacity market prices is the value of being able to sell power into higher priced external markets by withdrawing from the capacity market and eliminating the associated recall right. Because supply in the capacity market otherwise does not change in the same time frame in which retail access markets settle imbalances, the market price of capacity in a daily market, absent the exercise of market power, is likely to either be zero or rise to the level of the deficiency payment.

Moreover, if suppliers are able to withdraw from the capacity market to avoid the recall obligation, daily imbalance pricing in the capacity market can lead to supplier behavior that turns capacity market prices into a mirror of energy market prices (eliminating the smoothing role)[56] and lead to LSE behavior that undermines the reliability function of the capacity market. Under capacity market systems, LSEs that do not contract for sufficient generation to meet their capacity market requirement must pay a deficiency charge. If the price of capacity is volatile and varies day-by-day based on the value of capacity in external markets and LSEs have the option of paying daily deficiency charges instead of buying capacity credits, LSEs will have an incentive to buy capacity credits when the daily price is less than the deficiency charge and to pay the

---

56 While one could design a capacity market to have daily prices that are as volatile as those in an energy-only market, there is not much point to having a capacity market if the volatility that would otherwise be present in the energy market is simply shifted into the capacity market. NEPOOL initially had such a system in its operating capability market, which tended to track energy market conditions, and was eliminated after a relatively brief experience.

deficiency charge when the capacity credit price would be higher. This behavior can produce a shortage of capacity resources during precisely the days on which it is needed as suppliers pull out of the capacity market to avoid recall whenever the differential value of the capacity outside the capacity market exceeds the deficiency price.

The potential problems that a daily balancing market for capacity credits can give rise to are discussed in Section 2 below with reference to PJM. Section 3 then discusses how NYISO and ISO-NE have addressed the daily balancing problem.

2. Daily Balancing in PJM

In PJM, LSEs are not required to procure capacity credits to cover the loads they serve until the day before the operating day. In addition, capacity suppliers can pull their capacity in and out of the PJM capacity market (thus avoiding the recall obligation) on a day-to-day basis.[57] These market rules have tended to drive the PJM capacity market into a cycle in which the price is either zero or rises to the deficiency payment. On the one hand, the cost of keeping a unit available on a day-to-day basis is essentially zero because a supplier cannot mothball or activate units from day-to-day.[58] Thus, in the very short run the only cost of supplying capacity is the revenue foregone by being unable to sell non-recallable energy into adjacent markets.[59] What this means is that the value of PJM capacity is generally zero on a daily basis but occasionally spikes on those individual days when there is high demand for capacity both within and outside of PJM and the short-term value of capacity closely parallels the value of energy in external energy markets.

---

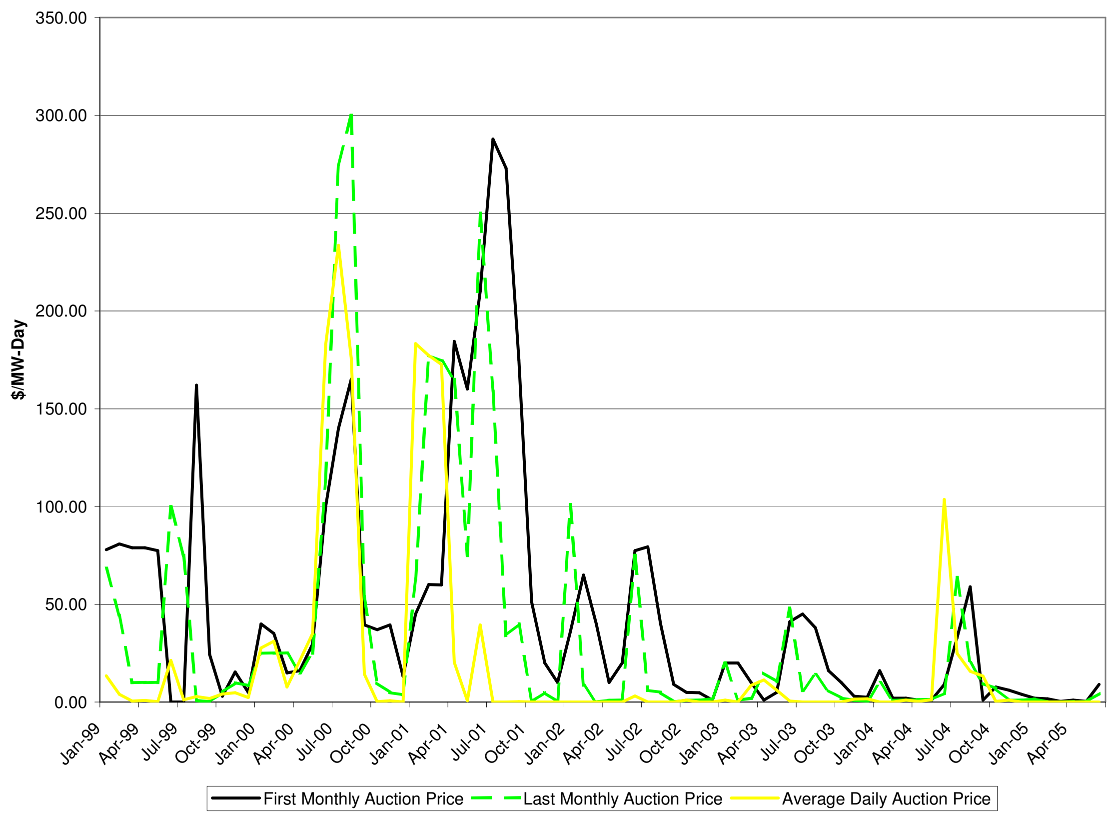
[57] PJM Operating Agreement, Schedule 11, Section 6.

[58] The costs referred to are the costs of having a unit available to be committed in the PJM day-ahead market, not the commitment costs themselves, which would be recovered either in energy prices or through the bid production cost guarantee.

[59] In PJM, as in NYISO and ISO-NE, resources that are committed to PJM as capacity resources are permitted to sell energy out of PJM under ordinary conditions, but any exports purchased from the PJM spot market or sourced from PJM capacity resources are subject to recall by PJM during shortage conditions, even if the price is much lower in PJM than in adjacent regions.

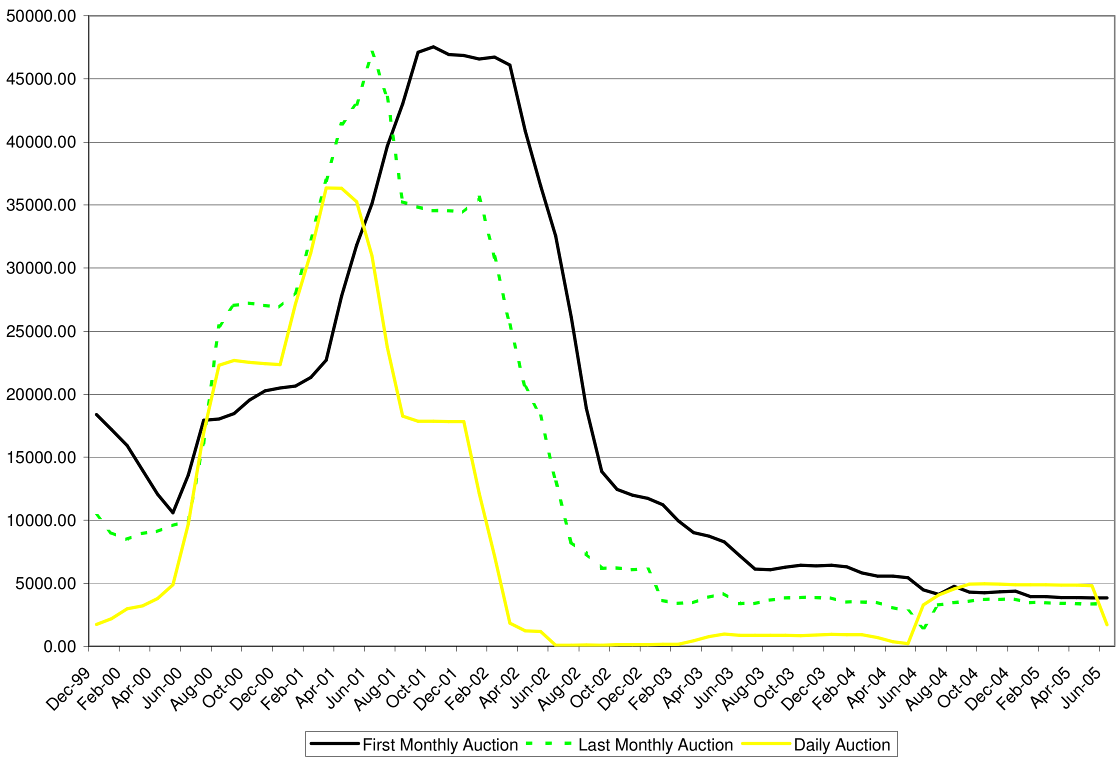
Thus, over the 1,073 days from January 1, 1999 through March 31, 2004 the PJM UCAP payment averaged less than $1/MWday. In fact, the PJM UCAP payment averaged less than $1/MWday in 32 of 63 months in that period. The trend is even worse with the UCAP payment averaging less than $1/MWday in 25 of the 33 months since June 2001. Conversely, however, the daily clearing price exceeded $160/MW for 73 days over the June through August 2000 period and 85 days over the January through March 2001 period. PJM also runs various monthly and strip auctions with similarly volatile prices. Figure 12 portrays the volatility of PJM UCAP prices.

Figure 12
PJM UCAP Market Prices



Overall, the prices in the daily UCAP market have yielded a annual per MW UCAP payment of slightly more than $7,000 over the period January 1999 through June 2005, but only about $1,500/MW year since June 2001. The prices in the first monthly auction covering each month averaged $16,700/MW year over the period since January 1999 and about $12,000/MW year since June 2001. Similarly, the prices in the final auction for each month averaged a little over $17,000/MW year since January 1999 but only a little more than $5,000/MW year since June 2001 (see Figure 13 below and Table 43, appended). Overall, PJM capacity market prices have been very low since mid-2001.

Figure 13
PJM 12-Month Rolling Average UCAP Capacity Payment
($/MW Year)



While these are very low payments compared to the long-run cost of capacity, the low payments since June 2001 can perhaps be viewed as the logical outcome of a capacity glut (except that, as noted above, the capacity glut is limited to certain subregions).

One limitation of the initial PJM capacity market system that contributed to this outcome was that PJM calculated deficiency payments on a daily basis, prorating the annual deficiency charge over the number of days the LSE was short. Thus, with a annual deficiency charge of $58,400, this prorated to a charge of $160/day for a deficiency (on a nominal capacity basis, when calculated in terms of expected capacity (UCAP) the deficiency charge was somewhat higher). If the going-forward cost of keeping capacity available in PJM worked out to $25,000/MW on an annual basis, this value might be recovered in two or three summer days when the value of capacity outside PJM ranged from $1000 to $15,000. An LSE could avoid these costs, however, by simply going short on those three days and paying the deficiency charge, but this left PJM short of capacity on the hottest days of the year when capacity was worth more outside PJM, leaving PJM unable to recall this capacity to maintain reliability.

PJM has attempted to address these incentive problems by developing rules that limit the ability of market participants to pay the daily deficiency charge to circumstances in which load shifts have resulted in the deficiency. Thus, effective July 1, 2001, if an LSE is short of capacity credits due to a load shift, it is assessed the deficiency penalty, prorated on a daily basis. LSEs have from 10-40 days to cure such a deficiency before PJM deems it not to have resulted from a load shift. Capacity market deficiencies not attributable to load shift are assessed the deficiency charge for the entire capacity market interval.[60]

3. New York/New England

In New York (and New England), each LSE is required to procure capacity in advance of each month based upon the loads the LSE is expected to serve at the beginning of that month. The NYISO capacity market consists of three auctions. The first is the capability period or strip auction which is conducted 30 or more days prior to the start of each six month capability period. The second is the monthly auction in which capacity is bought and sold for each individual month remaining within the current capability period. The third auction was originally called the deficiency auction, and is now called the spot market auction (under the current UCAP demand curve system).

There is no daily capacity market auction in New York or New England. Instead, New York and NEPOOL LSEs that lose loads during the month are credited for the value of the capacity acquired to serve the loads they lost. The ISOs calculate an adjusted daily capacity market requirement for each market participant as customers are gained and lost.[61] The value of that capacity is based upon the price paid in the ISO capacity market auction for that month, prorated for the part of the month in which the LSE did not serve that load. LSEs that gain loads are assessed a charge that is calculated in a similar manner. This avoids the indeterminancy of daily capacity market prices but capacity

---

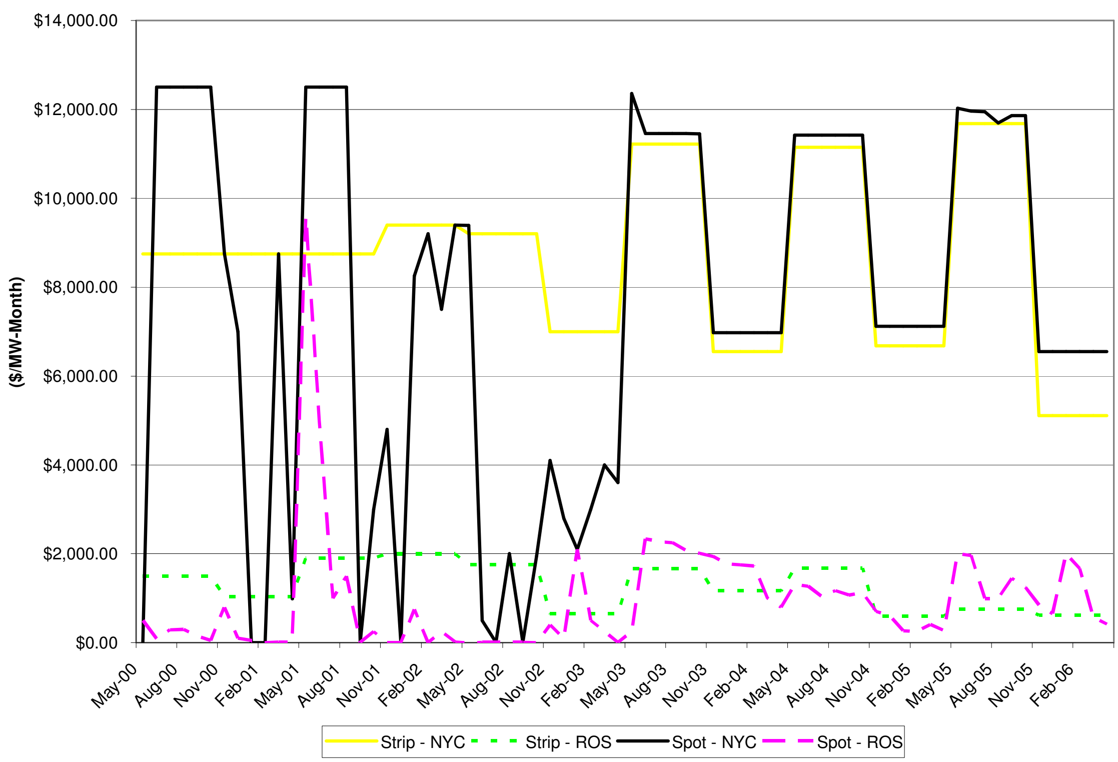
[60] Thus, if an LSE is deficient because it has gained load, it can pay the daily deficiency charge for a period of time. If an LSE becomes deficient because it did not buy enough capacity during a period in which its load remained constant or declined, then it is assigned deficiency charges covering the entire capacity market interval. This amounts to requiring the deficient LSE to pay the deficiency charge for every day of the interval on the largest deficiency it incurs for reasons other than load shifts on any day of the interval. PJM Reliability Assurance Agreement, Schedule 11, Sheets 53-54; PJM Unforced Capacity Market Business Rules, p. 5.

[61] NEPOOL Installed Capacity Manual; NYISO Installed Capacity Manual, Section 3.5.

market prices are also likely to be indeterminate on a monthly basis, because capacity also does not enter or leave the market from month to month.

New York initially considered a deficient LSE to be deficient for a six-month period. At present, New York purchases capacity to cover the obligations of LSEs that have not nominated sufficient capacity to cover their obligations for a month through a centrally conducted auction. LSEs can nominate resources to meet their share of the requirement, but if they do not do so, the ISO will buy capacity credits for them for that month in this auction (and send them the bill). Figure 14 shows that locational strip auction capacity prices in New York City have been relatively high, while rest-of-state capacity market prices have been relatively stable at around $22,000/MWyear.[62] The prices in the New York deficiency auction have been more variable and began in 2002-2003 to resemble the zero prices seen in PJM, despite the fact that New York does not have substantial excess capacity in its generation market.

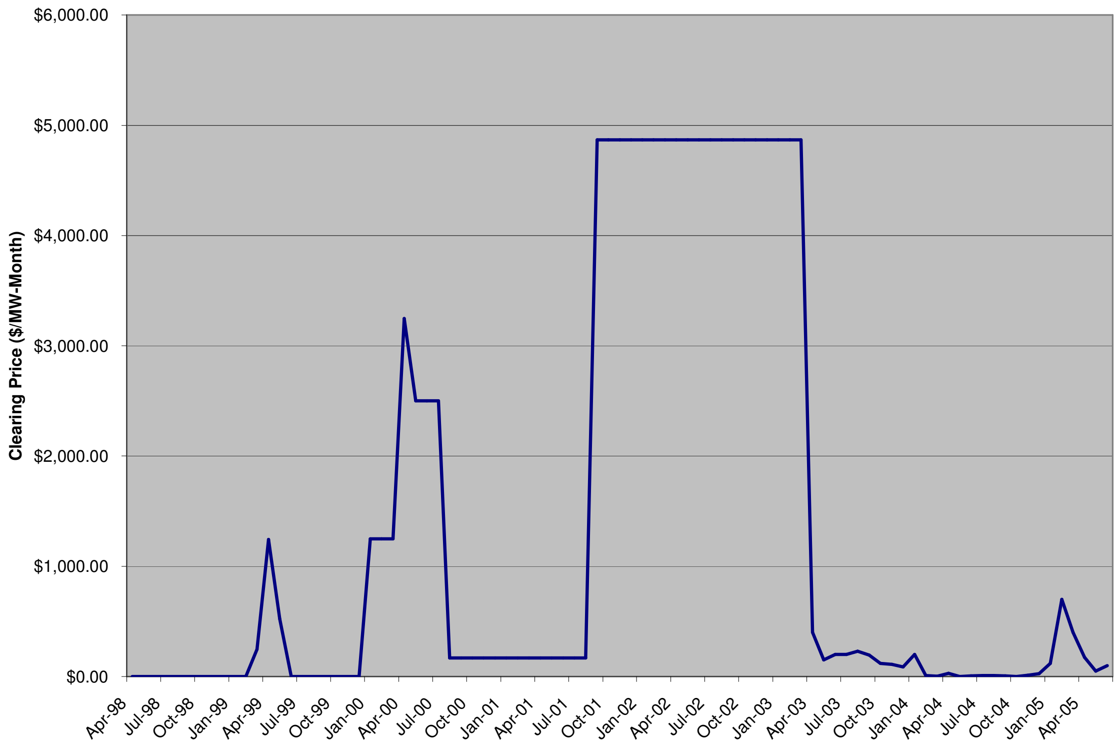
Figure 14
NY Capacity Market Auction Results



62 The prices tend to be lower in the winter because the NYISO calculates a single annual capacity requirement but uses summer ratings to determine summer supply and winter ratings to determine winter supply.

New England's capacity market has worked poorly. Initially, both the capacity requirement and capacity market prices were determined after the fact, making any kind of rational market behavior difficult to achieve. UCAP prices were generally zero through 1999, then rose substantially in the first seven months of 2000 (see Figure 15, below, and Table 42, appended).

Figure 15
NEPOOL Monthly Capacity Market Results



In early 2000, ISO-NE sought to terminate its capacity market on the basis that it was subject to manipulation, even though the capacity prices were quite low. This termination was approved by FERC and ISO-NE replaced the capacity market auction with a token deficiency charge and no capacity market auction. FERC did not approve this token deficiency charge and it was ultimately replaced with a $4,870 MW/month deficiency charge,[63] effective September 2001.

---

[63] It can be seen in Figure 15 that the FERC mandated capacity charge was higher than the capacity market prices alleged to have been impacted by market manipulation.

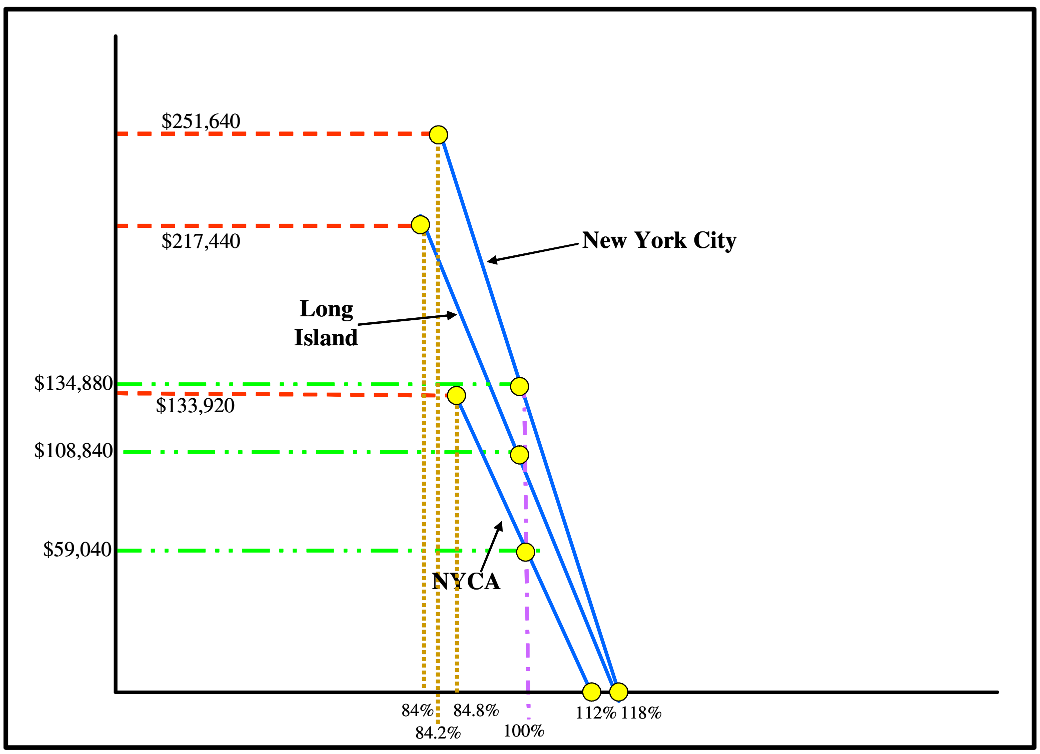
In April 2003, NEPOOL implemented an capacity market modeled after New York's, but capacity prices have been low, presumably reflecting the overall capacity surplus in New England.[64] New England's procedures were then similar to New York's system prior to implementation of the capacity market demand curve. The primary difference was in the way that the monthly auctions were conducted. ISO-NE simply offered to purchase the amount of capacity needed to cover obligations of LSEs that had not nominated sufficient capacity to cover their obligations. This amount was capped by the deficiency charge. In 2004, as noted above, ISO-NE filed to implement locational capacity market requirements and a capacity demand curve modeled after the New York capacity market.

[64] As in PJM, New England has an overall capacity surplus but does not have a surplus in all subregions.

4. *Capacity Market Demand Curves*

Beginning in 2003, the NYISO has defined a capacity market demand curve that is applied in the NYISO spot market capacity auction. Instead of being fixed, the amount of capacity purchased depends on the price of capacity. A target price is set for the target level of capacity, and if more than the target level of capacity is offered at the target price, more capacity is purchased and the capacity price falls. Conversely, if less than the target level of capacity is offered in the auction, the quantity purchased falls and the capacity price rises, but the price rises along the demand curve rather than the market price rising to the level of the deficiency charge. The initial demand curves for capacity for Summer 2003 are portrayed in Figure 16.[65] The target price figure for New York City, $127,890/MW year increased to $151,140 in 2004. Similarly, the initial Long Island capacity market target price escalated to $123,940 in 2004, while the NYCA target price escalated to $67,490. An independent study was to determine the capacity market target price beginning in 2005.[66]

Figure 16
NYISO Summer 2003 Capacity Demand Curve



65 Arthur Desell, Installed Capacity (ICAP), The Reliability Market.

66 NYISO Installed Capacity Manual, pp. 5-6 to 5-7.

The Summer 2004 capacity market demand curves are portrayed in Figure 17 and were relatively similar to the 2003 capacity market demand curves.

Figure 17
NYISO Summer 2004 Capacity Demand Curve

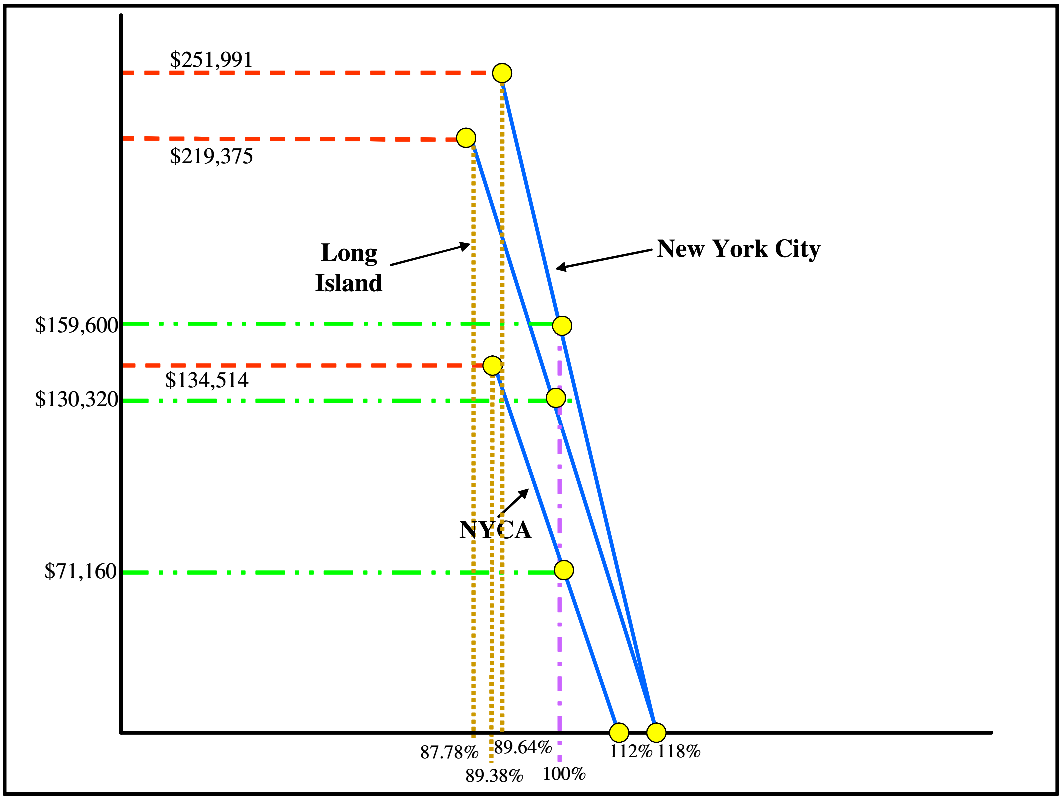
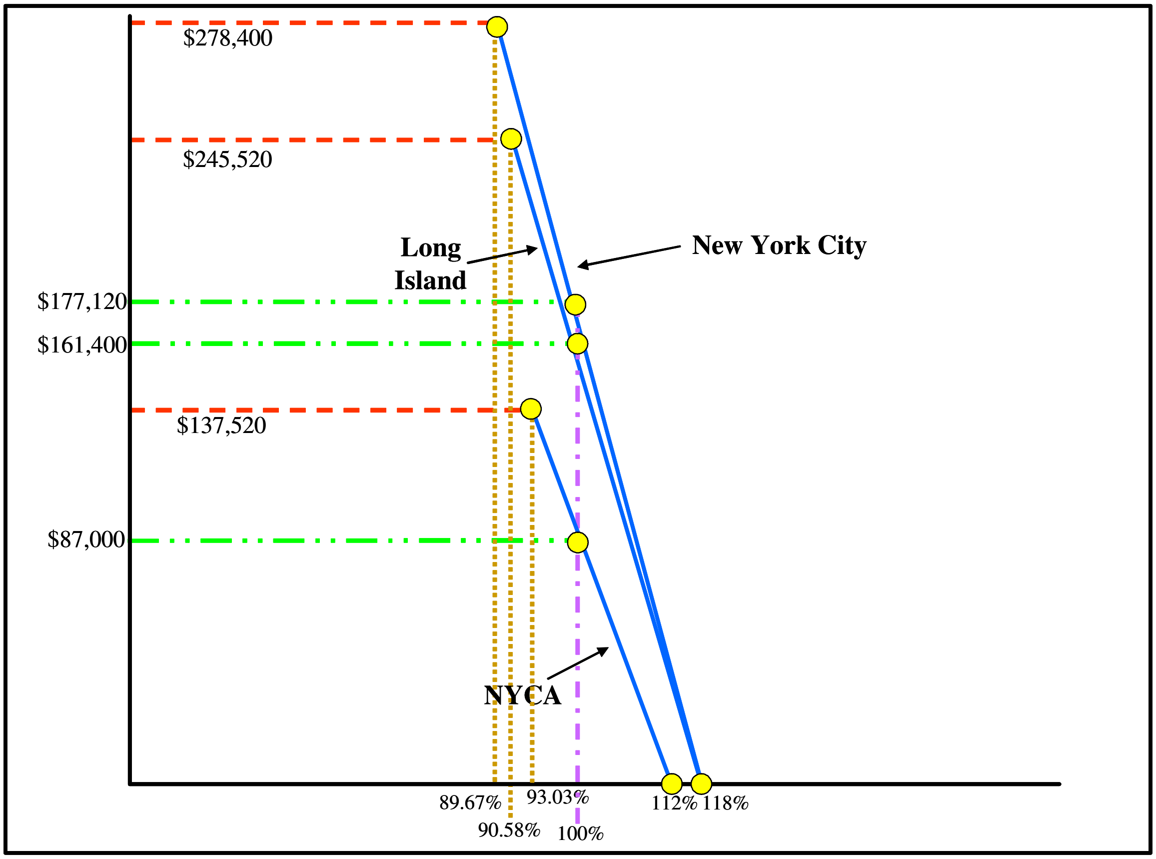


Figure 18 portrays the Summer 2005 capacity market demand curves. It can be seen that the target and maximum capacity market prices increased further.

Figure 18
NYISO Summer 2005 Capacity Demand Curve



With implementation of the NYISO capacity market demand curve, the NYISO spot market auction replaced the NYISO deficiency allocation. The capacity market demand curve is applied in the capacity spot market auction. Unforced capacity is offered into the capacity spot market auction to determine both the quantity of unforced capacity purchased in the auction and the price.[67] LSEs that have not purchased sufficient capacity to cover their obligation become net purchasers, while LSEs with excess capacity or resource suppliers are net sellers.[68]

---

67 Unforced capacity or UCAP is discussed in Section D.

68 See NYISO Services Tariff, Section 5.14. See also John Charlton, "NYISO Demand Curve as Proposed for the NYCA Installed Capacity Market," March 14, 2002.

As in New York, the ISO-NE demand curve proposed in early 2004 was to be implemented in a monthly spot market auction in which each participant would be required to offer all of their UCAP resources, with offers subject to mitigation. All capacity would have been required to be offered in the capacity region in which it was located and all participant load would be cleared in the capacity region in which it was located.[69]

The estimated full recovery capacity price used in determining the capacity market demand curve proposed by ISO-NE in early 2004 would have had two components. First, it would include the estimated price of new capacity of $6,666/MW month. Second, the estimated infra-marginal energy market revenue of a new GT, $2,100/MW month, would have been subtracted, yielding an capacity price of $4,566/MW month at the target capacity level.

The ISO-NE demand curve was originally targeted to recover the estimated cost of new capacity at 106.7 percent of the NEPOOL objective capability target, implying a reserve margin of roughly 18 percent over peak load, similar to the margin in New York. The capacity market demand curve was then to slope down to a zero payment at 118 percent of objective capability or roughly 112 percent of the full recovery price.[70] The demand curve was to be capped at 95 percent of objective capability, or roughly 89 percent of the full return capacity target.[71] The ISO-NE demand curve proposal was modified in the Fall 2004 to cap the demand curve at 100 percent rather than 95 percent of objective capability. In addition, a kink was introduced in the demand curve with the slope above the kink being three times the slope below the kink.[72] The details of the ISO-NE demand curve proposal evolved over time and many were the subject of dispute in an administrative proceeding. The focus of this paper is on the general concepts underlying capacity demand curves.

---

69 ISO-NE March 2004, pp. 42-44.

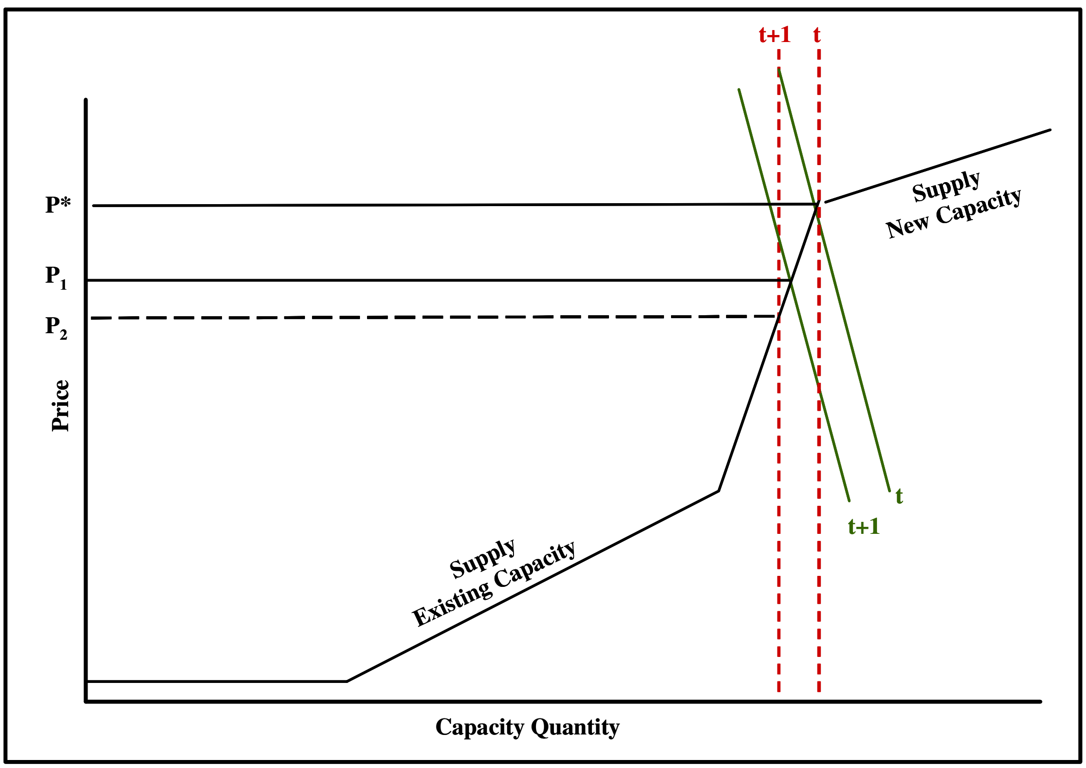
70 ISO-NE March 2004, pp. 19-21. Many changes were made in the ISO-NE proposal in August 2004 and in subsequent modifications of that proposal. This evolution is not covered in this paper.

71 ISO-NE March 2004, pp. 23-25

72 Stoft 2004, pp. 14-17, 75-80.

The effect of the capacity demand curve is to make the short-run demand for capacity somewhat elastic, which reduces the likelihood that a small capacity surplus will dramatically reduce the capacity market price. This short-run stabilizing influence is illustrated in Figure 19 which shows that variations in the capacity target relative to the level of existing capacity, such as the decline in the capacity target between t and t+1 in Figure 19, will lead to smaller variations in the price of capacity if the demand for capacity is downward sloping rather than vertical. Given the steepness of the short-run supply curve, introducing some slope into the demand curve results in relatively little short-run impact on the quantity of capacity procured but moderates the year to year variability in capacity prices.

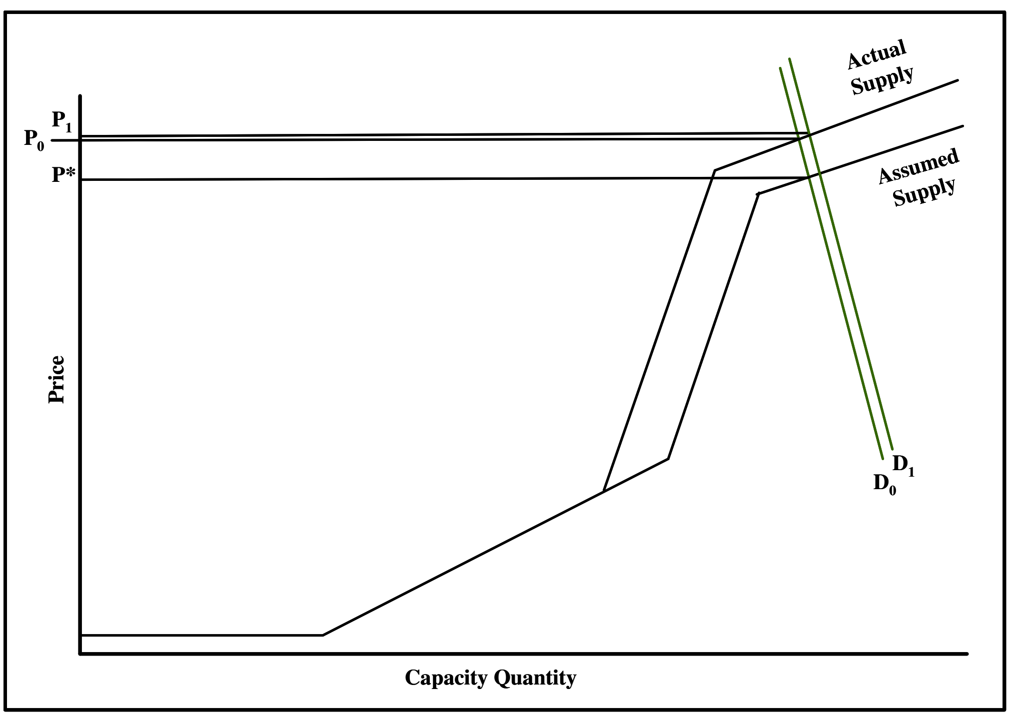
Figure 19
Impact of Peak Load Forecast Variations on Capacity Prices



In this circumstance a capacity demand curve provides more stability in the capacity payment to existing capacity than capacity market systems based on a fixed capacity requirement as the price of capacity would not fall as low during periods of surplus as under the fixed capacity requirement. It is apparent from Figure 19 that the price of capacity will be higher under the capacity demand curve model. While this higher price may appear to be disadvantageous to consumers, it needs to be kept in mind that in a competitive capacity market, the annual supply curve of new capacity depends on expectations regarding future capacity prices. If the expected level of post-entry capacity prices are higher, this would reduce the price increase required to induce new entry. If capacity investors were risk neutral and the capacity market perfectly competitive, the introduction of the demand curve would not materially change the price of capacity over the investment cycle, it would simply smooth the time path of expected capacity prices and new capacity would enter at lower initial capacity prices. In addition, the slope of the demand curve would cause customers to pay for somewhat more capacity than would otherwise be the case when capacity prices are low. Whether this is desirable, of course, depends on whether the price paid for that capacity is well related to the value of that incremental capacity.

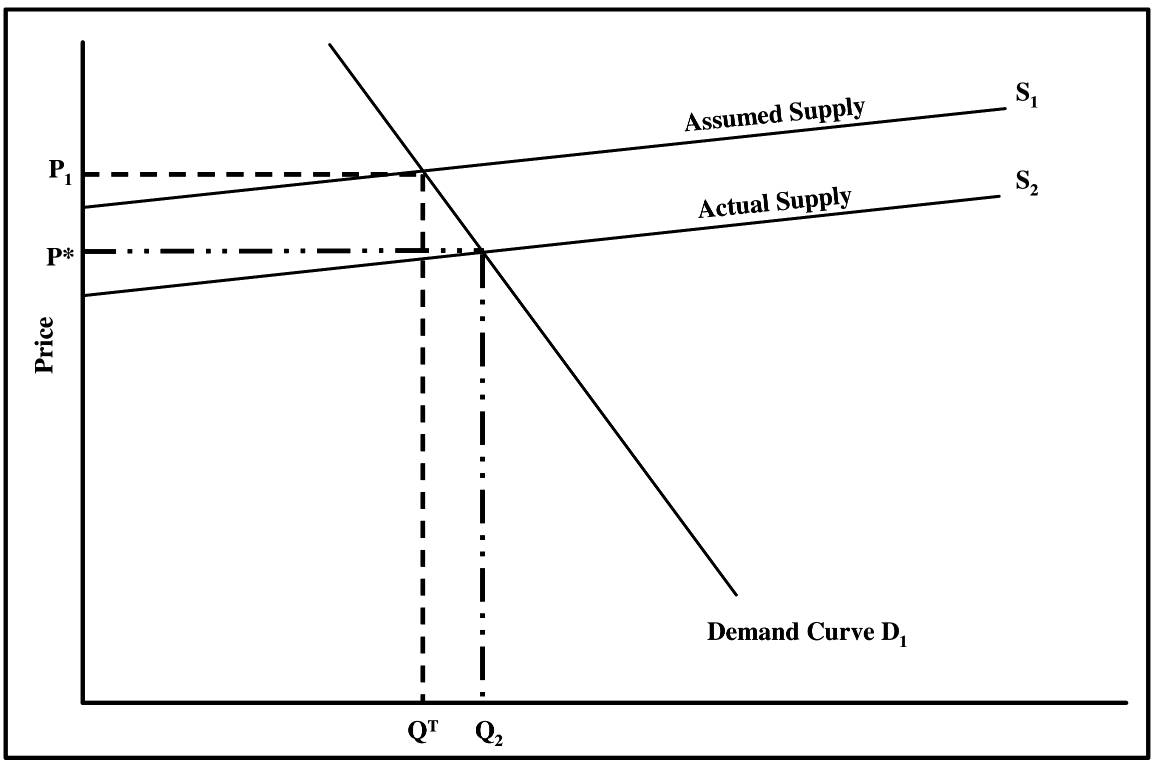
The development of the capacity demand curve, however, requires that the ISO estimate the market price of the target level of capacity and design the demand curve for capacity so that it intersects the supply curve for capacity at the target level of capacity. Errors in estimating the location of the market supply curve will result in variations in both the price of capacity and the actual level of capacity procured. If the capacity demand curve is relatively steep, an error in estimating the supply curve will have relatively little impact on the quantity of capacity acquired. The impact of the error on the price of capacity will be to raise the actual capacity price above the expected price (P^o compared to P*) but below the capacity price, given the mistaken expectations regarding supply costs, if the target level of capacity were purchased based on the actual supply curve (P_1), as seen in Figure 20.

Figure 20
Impact of Errors in Estimating Capacity Supply Curve



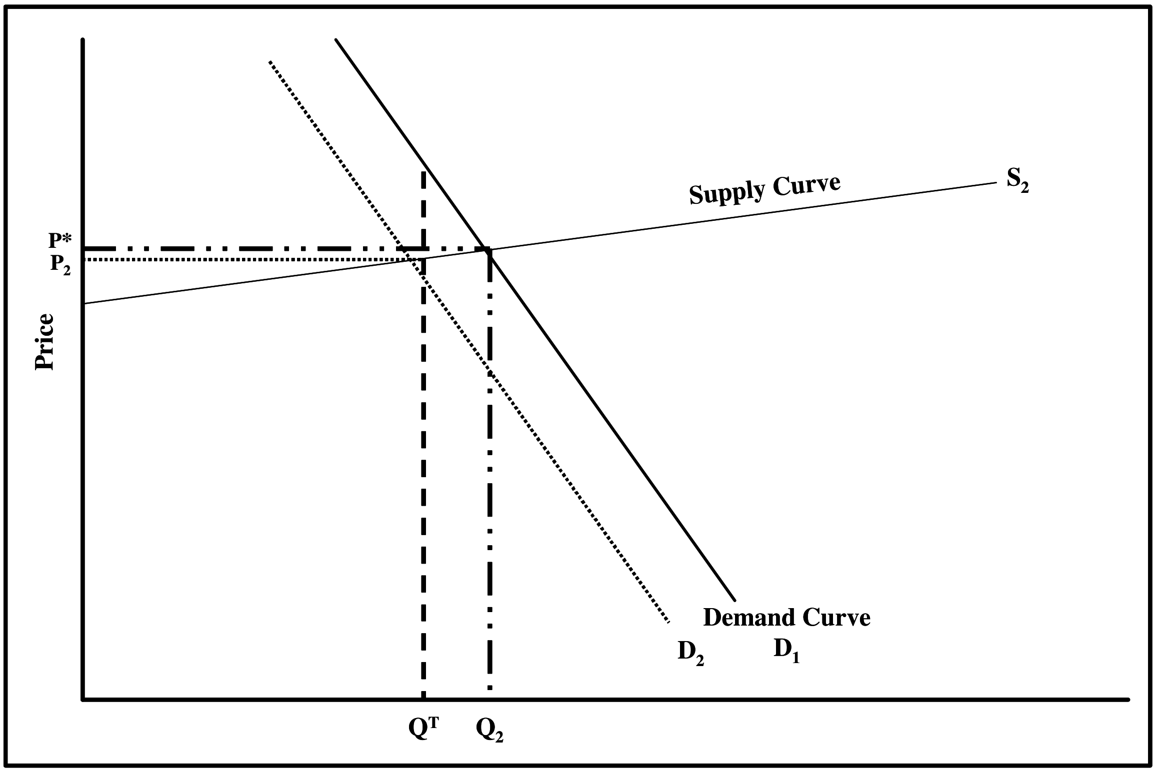
In the long run the supply of capacity is relatively elastic, so errors in estimating the location of the supply curve should lead to small variations in capacity prices and larger variations in the quantity procured. This outcome is illustrated in Figure 21. Suppose that S_1 is ISO's estimate of the capacity supply curve, and Qt is the target level of capacity. Given the ISO's estimate of the supply curve and its target capacity level, the ISO uses the capacity demand curve D_1 to clear the capacity market, expecting to observe price P_1. Suppose, however, that the actual long-run supply of capacity is described by the line S_2. Over time, the ISO will observe a level of capacity centered around Q_2 rather than Qt and a price of capacity centered around P* rather than P_1.

Figure 21
Long-Run Impact on Capacity Prices and Quantities of Misestimating Supply Cost



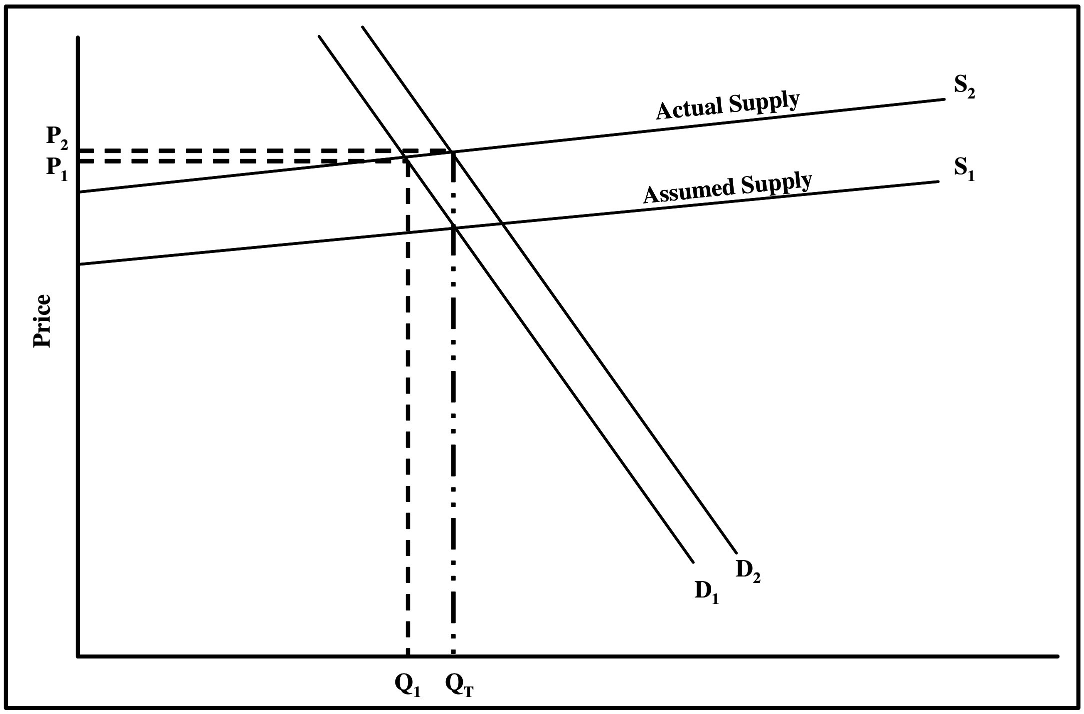
Had the ISO estimated the actual supply curve it would have used the demand curve D_2 to purchase capacity, and the price of capacity would have been centered around P_2 for the quantity Qt of capacity as shown in Figure 22. Overestimating the cost of capacity therefore raised the cost of capacity from P_2 to P* and would also cause the ISO to purchase excess capacity (Q_2-Qt) at a cost of (Q_t-Q_2)* P*. Not all of the cost of the excess capacity would be a social loss, as the additional capacity would make some contribution to improved reliability, but the contribution would be less than its cost.

Figure 22
Long-Run Impact of Misestimating Capacity Supply Costs



Conversely, if the ISO underestimated the cost of capacity, then the ISO would on average purchase too little capacity (Q_1 instead of the target, Q_T) at too low a price (P_1 instead of P_2) as illustrated in Figure 23. While consumers would benefit from the lower capacity costs, they would incur costs of reduced reliability that would more than offset the cost savings.

Figure 23
Impact of Underestimating Capacity Supply Costs



The error in the quantity of capacity purchased is the product of the error in the estimate of the supply price of the target level of capacity times the slope of the UCAP demand curve. Thus, the steeper the UCAP demand curve, the smaller the impact of errors in estimating the supply curve on the quantity of capacity purchased. The impact of the error on the price of capacity is then the product of the error in the quantity of capacity purchased times the slope of the supply curve, so the price impact will be smaller the flatter the supply curve.

The potential consequences of errors in estimating the cost of new capacity have been an important issue in debates over demand curve based capacity systems such as the ISO-NE LICAP system. At root these disagreements reflect the real difficulty in forecasting capacity costs, even for new capacity. Reasonable people can hold different expectations regarding the likely level of energy and reserve market earnings, the cost of acquiring and permitting a site and carrying out interconnection studies,[73] the economic life of capacity, the cost of capital, and expectations about the variability of capacity payments.[74] There is a truth or consequences aspect to the determination of the demand curve, however. If LSEs and regulators agree on the shape and location of the demand curve and it understates supply costs, this will result in adverse reliability impacts because too little capacity will be purchased. In a sense the debate may be best resolve by letting regulators and LSEs choose the shape of the demand curve and suffer the consequences.

This auction structure both reflects the belief of the NYISO and NYPSC that incremental capacity has some reliability value and a concern with eliminating the potential incentive of capacity buyers to engage in strategies that artificially depress the capacity price.[75] The capacity demand curve in New York appears to have stabilized capacity market prices in the spot market auction and ISO-NE included an ICAP demand curve in its initial proposals for a locational capacity market system.

A capacity demand curve system has several potential limitations. First, the market price of capacity is stabilized in part because the demand curve in effect attempts to guess the market-clearing capacity price, with the price rising or falling around this guess to balance supply and demand. If the estimated capacity price is too high, too much capacity will be purchased at too high a price. While it is not necessary that the estimated capacity price exactly correspond to the competitive price to achieve reasonable outcomes, material errors in setting the estimated capacity price will produce capacity shortfalls or excess capacity purchases. While a steep capacity demand curve minimizes the impact on capacity procurement of errors in estimating the supply curve of capacity, the steeper the demand curve, the less the difference in price volatility from a vertical demand curve.

Second, the capacity demand curve may tend to undermine forward contracting for capacity because LSEs that forward contract are not hedged for the cost of the additional capacity above the target that clears in the spot market auction. This would be an important limitation if substantial long-term forward contracts for capacity were being entered into but this does not appear to be the case in eastern capacity markets in states with retail access programs. It has been suggested that the demand curve has reduced

---

73 Keeping in mind that the expected returns must cover not only the costs of developing successful projects but must also cover costs not recovered in projects that were unable to go forward because of permitting or transmission issues.

74 The point being that if the cost of capacity is $50,000/MWh over a 20 year period but the capacity price will average $10,000 a year during 5 of those years as a result of recessions etc, then the expected return must exceed $50,000 in the remaining years in order to attract energy.

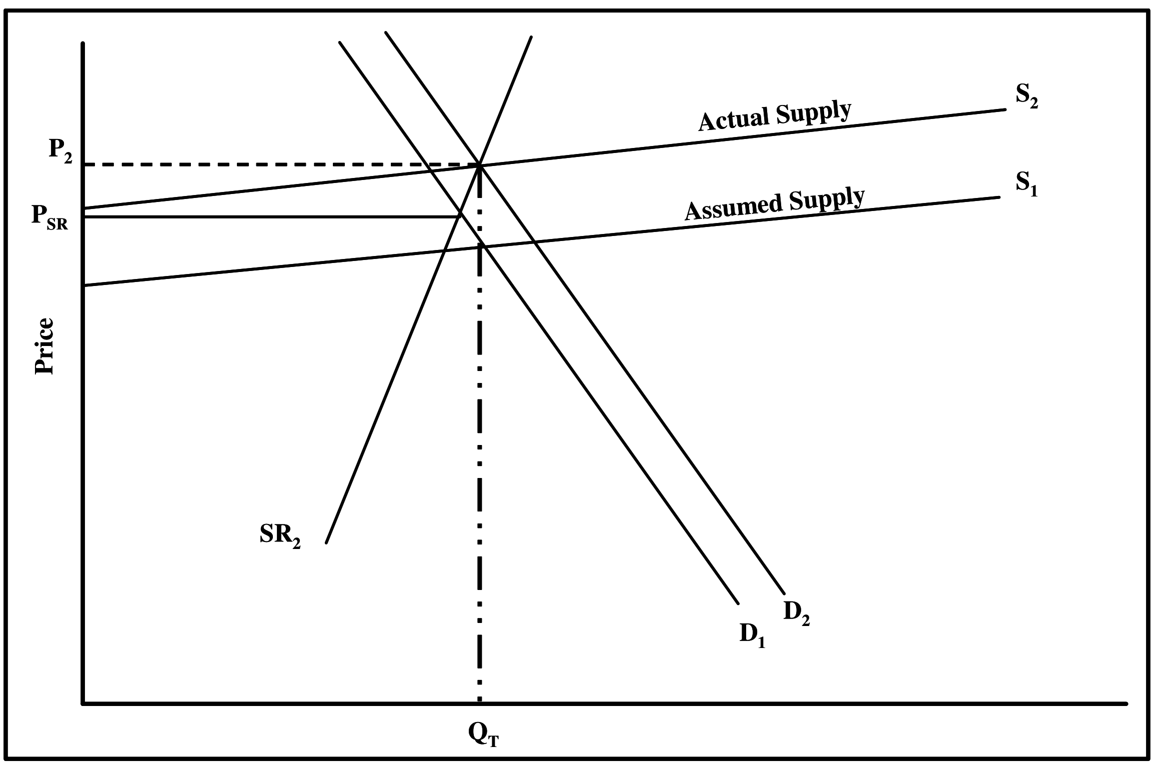
75 See Raj Addepalli, Harvey Arnott, Mark Reeder (New York PSC), Prepared Testimony Regarding a Proposal by the NYISO Concerning Electricity Capacity Pricing, March 6, 2003, pp. 3-5; New York PSC, Resource Demand Curve, January 31, 2003.

long-term capacity contracting in New York, but no data are publicly available. It appears that there has perhaps been a reduction in capacity sales in the strip auctions since implementation of the capacity demand curve.

The ISO-NE LICAP model, however, appeared to define the capacity target and the capacity demand curve in terms of megawatts of nominal capacity, without an adjustment for expected availability. This approach would give rise to potential reliability problems in using the demand curve to clear the market because the ISO would be purchasing a nominal amount of capacity rather than an expected amount of capacity. While low availability resources would only be paid in proportion to their availability during critical hours or shortage conditions, the amount of capacity the ISO needs to purchase in order to maintain reliability needs to be determined taking this expected availability into account. Thus, the ISO needs to use some measure of expected availability in clearing the capacity demand curve. A 50 MW resource with 50 percent expected availability should only count as 25 MW towards meeting the capacity target in the demand curve. The resource would then qualify for 50 MW of capacity payments but would be expected to only be paid 50 percent of the actual total because the resource would only be available during 50 percent of the shortage hours.

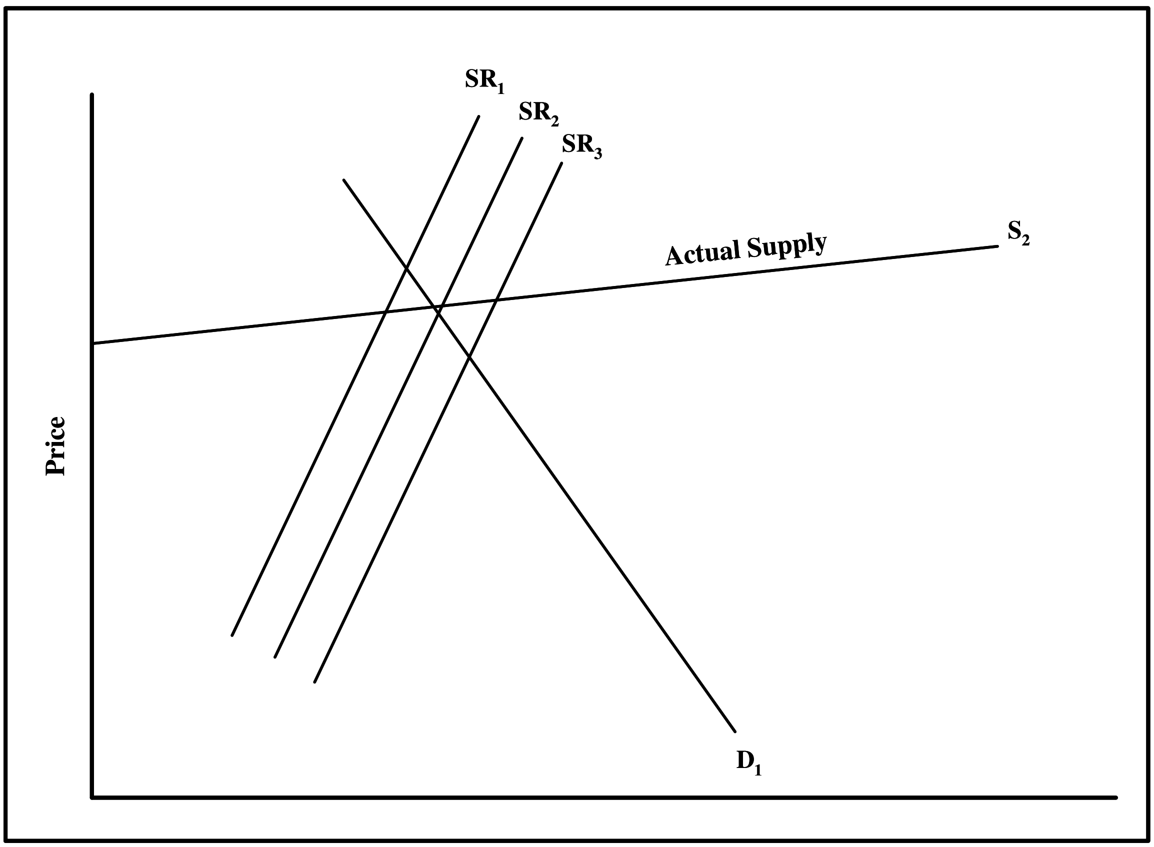
A third potential limitation of the demand curve approach is that while the impact of errors in estimating the long-run supply curve is small in the examples above in which the price of capacity is generally determined by the offer prices of new entrants along the long-run supply curve, as shown in Figure 23, the performance of a demand curve based capacity system may be less satisfactory if the price of capacity is generally determined by the offer prices along a short-term supply curve in which most supplier costs are sunk. Thus, suppose that suppliers expect the ISO to choose demand curve D_2 and they make long-run investment decisions in capacity accordingly, resulting in supply described by the short-run supply curve SR_2 in Figure 24. If the ISO selects demand curve D_1 for use in the auction, the capacity price will not be P_1 but P_{SR}, which is well below all the prices on the long-run supply curve.

Figure 24
Misestimated Supply Costs in Short-Term Capacity Markets



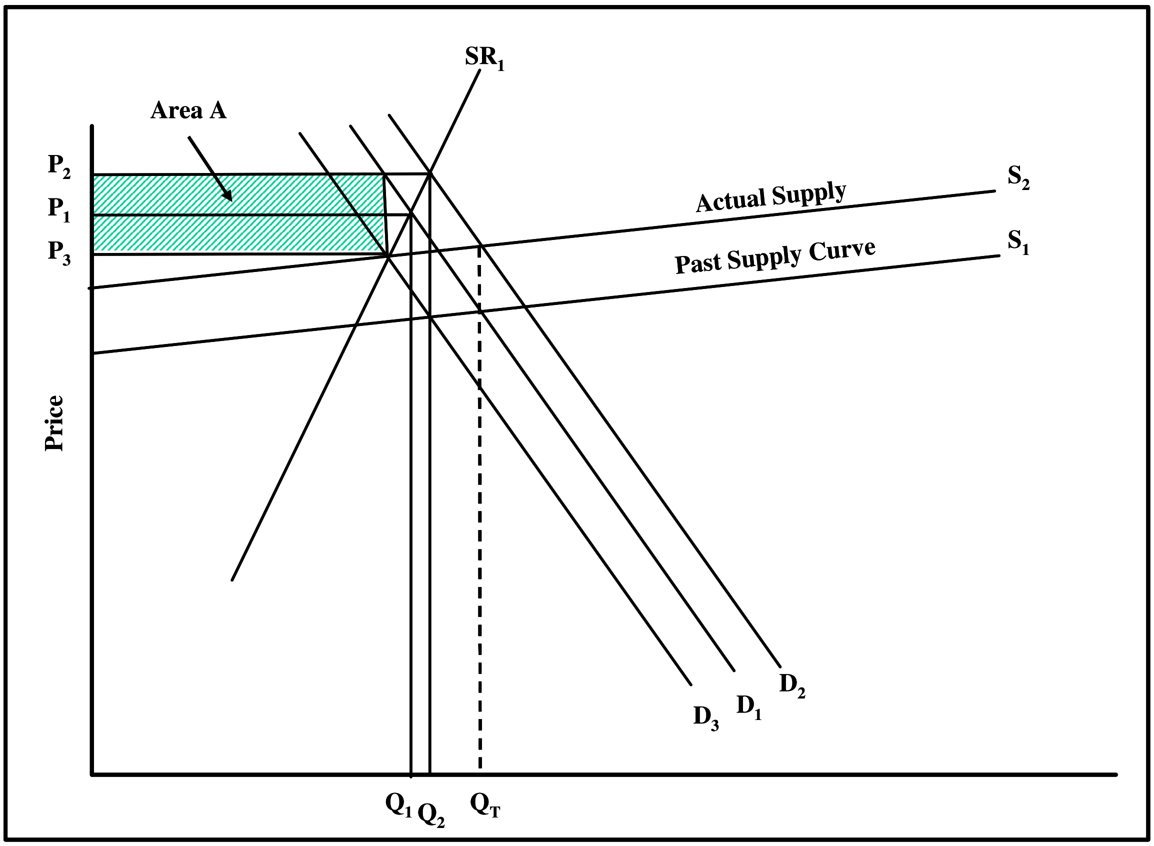
In a demand curve based capacity market in which the demand curve is determined and the auction held long after capacity investments must be made, the potential for errors, or biases, in estimating the long-run supply curve requires suppliers to base their investment decisions on forecasts of likely demand curves and prices. Given the kind of underestimated supply costs portrayed in Figure 24, suppliers might make investments resulting in short-run supply curves SR_1, SR_2 or SR_3 in Figure 25. Depending on which supply curve is offered, the price determined by demand curve D_1 could be above or below the actual long-run supply price.

Figure 25
Short-Run Supply Choices and Capacity Prices



Moreover, rather than "forecasting errors" working themselves out in such a model, they may become worse over time. When it is time for the ISO and regulators to set the capacity target and demand curve, the short-run supply of capacity will be pretty much determined and it will be apparent to those choosing the demand curve whether the short-run supply curve will be SR_1, SR_2 or SR_3. Suppose that it is apparent that the short-run supply curve in the auction will be SR_1. If the ISO were to stick to its past assumption that the long-run supply curve is S_1 and choose demand curve D_1, this would result in price P_1, with a short-fall of capacity relative to the reliability target of Q_t-Q_1. Alternatively, suppose the ISO recognized that it had underestimated the long-run supply curve in the past and wanted to move back to the capacity reliability target of Q_t by choosing demand curve D_2. This would result in a very large increase in the price of capacity to P_2 as illustrated in Figure 26. The magnitude of the resulting increase in the capacity price might be impossible from a political standpoint. Worse, there would be a temptation to choose an even lower demand curve, such as D_3, which would result in a slight decrease in capacity and a huge reduction in per-MW capacity payments (Area A).

Figure 26
The Long-Run Impact of Underestimating Supply Costs in a Short-Run Capacity Market



These kinds of problems would be particularly acute in a monthly capacity auction such as proposed in ISO-New England's LICAP model but are potentially present in any near-term auction, including the current NYISO market design.

5. Forward Capacity Markets

Another approach to addressing the short-term focus of capacity contracting has been to move forward in time the point at which the capacity market is cleared and capacity prices determined. Contracting for capacity three, four or five years prior to the operating year would likely smooth the time path of capacity prices, because there would be fewer gluts or surpluses of capacity that far forward. Such a multi-year forward capacity procurement process would tend to cause the demand for capacity, whether vertical or sloping, to clear along the long-run demand curve, rather than on the short-run supply curve, addressing the kind of problems discussed in the section above. A multi-year forward procurement process would also reduce the potential for the exercise of market power, because there would be sufficient time for more new capacity to enter in response to high prices.[76] This concept of forward contracts for capacity was an important element of NERA's CRAM proposal,[77] ISO-NE's recent FCM capacity model (FCM standing for forward capacity market)[78] and is an element of many of the call contract proposals.

Of course, there is nothing in any of the other capacity market designs that would prevent LSEs from contracting for capacity four or five years in advance, thus avoiding the boom bust cycle and addressing market power concerns. The PJM and NYISO capacity markets were originally developed envisioning that kind of behavior by the distribution companies comprising the pools. A feature of the current market design, however, is that the introduction of retail access has meant that very few load serving entities in PJM, New York or many parts of New England, appear to have an interest in entering into long-term capacity contracts, because they have no long-term load-serving obligations.[79]

The various forward capacity market designs address this problem by proposing that the system operator would run a forward capacity auction in which suppliers would submit offers several years prior to the operating year. Existing suppliers would generally sell their capacity rights one year at a time, several years forward, while some proposals, such as ISO-NE's FCM proposal, would allow entrants to lock in the capacity payment for several years. The demand for capacity might be either sloping or vertical depending on the details of the market design. Accepted offers would become binding capacity contracts whose costs would be assigned to the appropriate LSEs when the operating year arrived. Resources would be contracted to supply capacity for the entire capacity market year, with the changes allocated to LSEs on a monthly basis, which

---

76 Contracting for all capacity three years forward does not completely eliminate the potential for the exercise of market power. If there are asymmetries in the cost for incumbents and entrants within particular load pockets, the price of capacity would not be effectively constrained by the threat of entry. In addition, while a variety of projects are in development at any point in time, large increases in supply would not be forthcoming at the competitive price in the short-term. Entry would most effectively constrain prices if consumers were only contracting for a portion of their load three years out, with a significant proportion hedged under even longer term contracts.

77 Eugene Meehan, Chantale LaCasse, Philip Kalmus and Bernard Neenan, "Central Resource Adequacy Markets," Final Report, February 2003.

78 Settlement Agreement Resolving All Issues, Item 11, Section I (hereafter Settlement Agreement); Explanatory Statement, Section VI.B.3. See also Cramton and Stoft 2006, p. 16.

79 LIPA and other municipals are an important exception.

would tend to increase the likelihood of the capacity price clearing somewhere other than close to zero or at the deficiency payment.

LSEs that either own supply resources or that are willing to contract for capacity even further out than the ISO-coordinated forward capacity market are able to self-provide capacity to meet their obligations in the auction, in effect opting out of the ISO-coordinated process.[80] LSEs can therefore enter into their own capacity contracts and then offer this capacity in the capacity market auction to cover themselves against their future capacity credit obligations. In effect, the ISO would only be buying capacity for those LSEs who do not buy capacity forward on their own. If munis and coops, for example, bought some capacity seven years out and some capacity four years out, they might achieve lower costs than retail access customers without being adversely impacted by the FCM auction.

There are complications in the ISO coordinating such a forward auction, in particular the need to ensure that supply offers reflect resources that will be available with reasonably high probability in the operating year, without imposing rules that raise the cost of capacity by adversely impacting the ability of one or more sources of capacity to compete in the forward auction.[81] One approach to ensuring performance would be financial. For example, the ISO-NE FCM proposal provides for collateral payments by capacity suppliers cleared in the forward market to ensure that the bids reflect real capacity.[82] The other approach is to specify various eligibility requirements,[83] which have the potential to be asymmetric between resources. A particular example of such a potential asymmetry concerns demand response resources, which do not require three years to put in place and may be hard to contract for three years in advance with specific end use customers, while the demand response provider might have much less

---

80 Settlement Agreement, Item 11, Section II.F; Section III.D and Section V.B.3. As discussed below, loads self-providing capacity would also be able to in effect opt out of the peak energy rental adjustment, avoiding those potential distortions. There are a few ambiguities as to how capacity costs are allocated that could affect the economics of self-supply. It appears that under the FCM proposal self-supply offers from new capacity cannot lock in the auction price for five years, but conversely that none of the potential excess between the five year in advance price and the three-year forward auction price would be allocated to self-supplied load, but this is not clearly spelled out. In addition, if the ISO's three year out load forecast is high, the capacity costs in the operating year would exceed the load requiring capacity. It is not clear how these costs would be allocated between loads self-providing capacity and loads buying capacity through the auction under the FCM proposal.

81 The FCM proposal also provides for a variety of milestones that must be satisfied by new capacity resources. Settlement Agreement, Item 11, Section III.B.3.

82 Settlement Agreement, Item 11, Section II.G.2.

83 See, for example, Settlement Agreement, Item 11, Section II.A.3.

uncertainty in its ultimate ability to deliver the specified quantity of demand response capacity than a power plant developer that has not yet obtained all its permits.[84]

The ISO serving as in effect a backup forward procurer of capacity in a forward market may be an unqualified improvement relative to six month in advance capacity procurement mechanisms but is not necessarily preferable to addressing the design issues in retail access programs so that LSEs have appropriate incentives to contract forward for energy and capacity.

A potentially important limitation of these capacity market designs is that the ISO will be contracting forward for capacity to meet a three-year forward load forecast without regard to the price of power to consumers and whether the consumers for which the capacity is intended would actually demand power at those prices. This topic is discussed further in Section G below.

6. Discussion

The underlying problem that retail access poses for capacity market systems is that the provider of last resort and price to beat provisions of many retail access systems are fundamentally inconsistent with long-term commitments by LSEs. The resulting unwillingness of LSEs to make long-term commitments undercuts reliance on the capacity market to support entry and reliance on competitive entry to keep capacity markets competitive. Retail access programs pose similar problems for energy only markets so the problem cannot be addressed by switching to reliance on energy-only markets to maintain reliability.

Residential customers are unlikely to be willing to sign long-term energy contracts that lock in payments for capacity credits over a 5- to 10-year period. Given the typical rate at which people change houses, residential customers signing 5- to 10-year power contracts could be faced with buying out uneconomic contracts or trying to capture the value of in the money contracts from the new owner or renter. For whatever reason, residential contract duration is generally one year or less, too short to support long-term capacity market or energy purchase contracts.

Long-term contracts under retail access would likely pose fewer problems for commercial and industrial accounts. Long-term contracts that matched the willingness to pay of industrial and consumer customers to the cost of power would be helpful in avoiding future stranded costs. The recent actions of the New York Public Service Commission in ordering hourly pricing for large industrial and commercial load address

---

[84] The ISO-NE FCM proposal does not contain provisions concerning capacity provided by intermittent and demand response resources but provides that these rules are to be developed by the end of 4th Quarter 2006. Settlement Agreement, Item 11, Section II.E.

the underlying incentive problem by exposing these customers to real-time energy prices on a permanent basis.[85] This program should provide the incentive for more substantial forward contracting for energy and capacity and the results of this order should provide valuable information on contracting incentives.

Short-term sales contracts with residential consumers could nevertheless in principle support long-term fixed price capacity or energy purchase contracts by LSEs. Oil companies, for example, have supported the construction of refined product pipelines through long-term take-or-pay contracts with the pipeline, despite the fact that the refiners have no contract with retail motorists. While long-term forward fixed price capacity or power purchase contracts would increase the risk of retail access providers, the riskiness of these contracts could be managed by periodically entering into long-term contracts for only a portion of the retailers customer demand. Such a retailer would lose money when the market price of capacity or power fell below its long-term contract price, but it would make money when the market price of capacity or power rose above the long-term contract price. By entering into a temporally diversified set of power purchase contracts, such a retailer could limit its risk exposure to sudden swings in capacity prices.

It is not clear at present whether longer-term contracts between retailers and end-use customers would lead to a more efficient market equilibrium or would simply reduce average prices to consumers but shift corresponding risk from suppliers to consumers. Long-term forward contracting is only necessary if it produces efficiencies, either by eliminating market distortions or by shifting risk to entities with a lower cost for bearing risk. At present, there is limited empirical information bearing on this question in retail electric power markets. Indeed, one of the problems in analyzing these issues is the very limited data that are available regarding the forward contract decisions of either regulated or unregulated suppliers. Nevertheless, public statements by unregulated retailers indicate that these entities do not enter into forward contracts for energy or capacity for more than a year or so.

There are a number of possible reasons for this observed pattern, and more than one may be applicable in some cases. First, it is possible that there may not be any efficiency gains from forward contracting. If forward contracts would merely shift risk between entities more or less equally able to bear it, there would be no gain from improved risk bearing. While long-term contracts can provide more competitive pressure from entry on incumbents, this advantage would be minimal if spot price mitigation assured consumers of prices at or below the long-run competitive level without the need for forward contracts.

---

85 State of New York Public Service Commission, Case 03-E-0641, Proceeding on Motion of the Commission Regarding Expected Implementation of Mandatory Hourly Pricing for Commodity Service, September 23, 2005 and April 24, 2006.

Second, it is possible, particularly in Nepool and PJM, that the current capacity market designs and policies have resulted in spot capacity prices that are well below the going-forward costs of incumbent suppliers. Suppliers would be unwilling to enter into long-term contracts at capacity prices that fail to cover their going-forward costs, opting instead to shut down unless supported with RMR contracts. LSEs, on the other hand, would have no incentive to enter into long-term capacity contracts at the level required to cover resource going-forward costs if capacity can be purchased in the spot market at much lower prices. While this explanation could account for the apparent contracting problems in Nepool and PJM, it is not clear it would explain a lack of forward contracting in Zone J in the NYISO (New York City), where capacity prices are substantial. In addition, this consideration could account for a lack of interest in long-term capacity contracts by competitive retailers but would not account for a similar lack of interest in long-term congestion hedges.

There are also several regulatory features of retail access markets that may undercut long-term capacity or energy contracts that are not hedged by long-term customer contracts. The third reason for a lack of forward capacity contracts could be that such long-term contracts would only be economic if the losses incurred when the market price of capacity is below the contract price were offset by profits when the market price of capacity is above the contract price. If the retailer were a regulated utility that cannot retain such a difference between the contract price and the market price, then the risk of loss with no offsetting possibility of gain would preclude entering into either long-term capacity or energy contracts.

Fourth, unregulated LSEs may be unable to benefit from low contractual capacity purchase prices during periods of high capacity market prices if the regulated price to beat does not rise commensurately. Retail rate regulation policies that tend to keep the price to beat too high when capacity and energy prices are low and too low when capacity and energy prices are high would tend to discourage long-term capacity and energy contracts. With such a retail price structure it is more profitable for unregulated LSEs to shed customers back to the utility when capacity prices rise than to enter into long-term contracts that hedge capacity costs. This has unfortunately been a feature of a number of retail access markets; unregulated LSEs have a free hedge against increases in wholesale energy or capacity prices through their ability to send retail customers back to the provider of last resort if prices move unfavorably.

Finally, a fifth factor is the risk of regulatory change. Capacity credits are an artificial product. LSEs may be deterred from entering into long-term capacity contracts by the risk of regulatory changes. Particularly problematic would be regulatory change which retains the capacity market system, and thus does not trigger clauses terminating payments if the capacity market requirement is terminated, but which dramatically reduces the market price of capacity. ISO NE's decision to dramatically reduce the capacity market deficiency payment without eliminating the requirement would be an example of this kind of risk.[86] Another example of regulatory risk is the introduction of a locational capacity market in New England. A Boston LSE that had entered into a 10-year capacity contract with a resource in Maine would find itself no longer fully hedged.

Capacity market systems would be more workable in combination with retail access programs in which residential customer demand is covered with multi-year contracts as these contracts could provide the basis for multi-year capacity contracts. Still, the interval between the signing of the contract and the duration of the contracts limits the ability of the LSE winning the retail contact to contract with new generation entrants to meet its capacity market obligations. Unless the generation entrant is very far along in the development/construction process, it would not be on line in time to hedge the load contract.

D. Outage Performance

A second performance issue arising under capacity market resource adequacy systems is that rules are needed to ensure that the capacity receiving capacity payments is sufficiently reliable that it is available to meet load under stressed system conditions. As with capacity market deliverability requirements, outage standards are necessary because generators receive capacity market revenues whether or not they actually operate, so an additional mechanism is necessary to ensure that generators receiving capacity market payments have incentives to minimize forced and maintenance outages. This incentive problem has initially been addressed by the development of the UCAP system which is currently applied throughout the Northeast.

UCAP systems calculate a capacity requirement based on projected generation outages based on historical performance of capacity resources. The amount of capacity each supplier is entitled to sell is then determined by scaling down the nominal rating of the resource based on the resource's historical forced outage performance. The scaled-down capacity is called UCAP (in essence the expected capacity of the resource), and it is UCAP that LSEs are required to purchase. Under existing UCAP systems, the UCAP capacity of each supplier is fixed prior to each auction based on the forced outage performance of that supplier's resources during a prior period. ISO-NE calculates UCAP ratings monthly based on a rolling 12-month historical performance.[87] PJM calculates UCAP ratings for each capacity market interval.[88] The PJM UCAP rating is calculated on a rolling 12-month average for the 12 months ending 2 months prior to the billing

---

[86] Of course, there is also the risk of tightened requirements and FERC in fact ultimately restored a substantial deficiency payment.

[87] NEPOOL Manual for Installed Capacity, p. 3-10, January 1, 2004; New England Power Pool, Market Rule 1, Section 8.3.6, Sheet 87.

[88] The PJM capacity market intervals are June 1-September 31, October 1- December 31, and January 1 to May 31.

interval.[89] The NYISO calculates UCAP based on the average EFORd calculated for the six most recent 12-month rolling average periods.[90] NYISO UCAP is calculated separately for generation in New York City, Long Island, and the rest of state, so locational UCAP requirements are established. Under all three systems,

$$\text{UCAP} = \text{Nominal Capacity} * (1 - \text{EFOR}_d)$$

A significant feature of these UCAP systems is that the capacity payments to suppliers are based on the amount of UCAP supplied, effectively the expected value of the amount of capacity supplied, rather than the physical capacity. Units with poor historical forced outage performance are therefore able to sell less UCAP per megawatt of physical capacity and will earn less money in the capacity market in the future, motivating resources to maintain high levels of availability. In addition, this payment mechanism means that the total capacity payments by consumers are equal to the price of UCAP times the amount of UCAP purchased (i.e., the expected value of capacity) and do not depend on the underlying nominal physical capacity.

Importantly, the UCAP rating only reflects resource forced outages, which are assumed to be random, and does not reflect planned maintenance outages. The determination of pool capacity requirements assumes that planned maintenance outages (including the refueling outages of nuclear plants) occur during low load periods of the year in which the capacity is unlikely to be needed. As a result, capacity market systems need to ensure that planned maintenance outages are, in practice, scheduled in a manner consistent with the assumptions used in developing the capacity requirement.

Capacity market systems therefore restrict the scheduling of planned maintenance outages by capacity market resources. The NYISO requires that capacity market resources provide the NYISO with advance notification of outages and outages are subject to being rescheduled by the NYISO.[91] PJM requires that capacity market resources submit schedules of planned outages to PJM for coordination with other generation and transmission outages.[92] In addition, PJM deducts capacity that is unavailable due to maintenance during PJM's peak season (roughly mid-June to mid-September) from the resources' unforced capacity.[93] ISO-NE also required that capacity

---

89 PJM Manual, Capacity Obligations, Section 1, p. 6, June 1, 2005; PJM Reliability Assurance Agreement, May 17, 2004, Sheets 15, Schedule 5.1, Sheets 42-43.

90 NYISO ICAP Manual, Section 4.5, p. 4-12; NYISO Services Tariff, Section 5.12.6(a), Sheets 135B-135B.01.

91 NYISO ICAP Manual, Section 4.3.

92 PJM Reliability Assurance Agreement, May 17, 2004, Section 9.2.

93 PJM Reliability Assurance Agreement, May 17, 2004, Schedule 8.

market resources notify the ISO in advance of their proposed maintenance outage schedules and these outages are subject to being rescheduled.[94]

One negative side effect of UCAP systems is that generators appear reluctant to declare forced outages because of the impact of outages on their capacity market revenues. Instead, they may drag on the system when capacity having operating problems is dispatched to meet load rather than admitting to a derating. It is therefore desirable, in combination with such an UCAP system, to either have significant penalties for failing to follow dispatch instructions or some other system of sanctions. Outages and deratings that occur on a high load day are unfortunate from a reliability standpoint regardless of the market design but their reliability impact is exacerbated if the system operator is not informed of them until the units are unable to perform as instructed.

Moreover, since the EFORd forced outage data are employed in the Monte Carlo analysis used to determine capacity requirements, the incentive of generators to overstate availability can potentially impact reliability by leading to understated capacity requirements. NYISO audits identified such overstated unit availability in the GADs data supplied by some generation owners, leading to a material increase in the NYISO capacity requirement when the higher EFORd outage rate was used in the Monte Carlo reliability analysis.[95]

A second issue is whether the EFORd index employed by UCAP systems to measure generator availability provides sufficient performance incentives for baseload units.[96] The EFORd UCAP systems employed in the Northeast essentially cause a capacity market supplier to receive the capacity payment in proportion to its availability. Thus, if capacity market resource were on line 6,650 hours, and out due to forced outage in 350 hours, it would have a 95 percent EFORd rating and would be paid for 95 percent of its capacity (or one can think of this unit as being paid 95 percent of the capacity price). This would be the case whether the 350 hours of forced outage occurred in the spring when the price of power was $12/MWh and the outage had no reliability impact or if the 350 hours of forced outage occurred in July, when the average LMP price was $500 and the outage resulted in load shedding.

Similarly, the incremental value of staying on line over a day is relatively small under a UCAP system. For a unit with around 7,000 hours combined on line and out of service, the impact on the unit's EFORd of a 24-hour forced outage would be a little

---

94 ISO-NE ICAP Manual, Section 3.3.

95 NYSRC 2004, p. 4, 23.

96 A related issue which does not appear to be a problem would be a potential for market participants to simply not offer capacity in real-time without declaring a forced outage. The ISO rules appear to deter such behavior. ISO New England market rules call for imposing a sanction of an amount up to the deficiency charge and imposing a financial sanction equal to the corresponding real-time LMP price. New England Power Pool Market Rule 1, Appendix B, p. 307.

more than .3 percent, so would cost a little less than $350/MW for a New York City unit with a $100,000/MW UCAP price. The UCAP system by itself therefore provides baseload units with relatively little incentive to make themselves available under stressed market conditions. This is not a problem if availability is purely a function of mechanical forced outages that occur randomly and cannot be influenced by supplier choices as supplier performance would be consistent with the assumptions in the Monte Cristo analysis that determined the capacity requirement. In fact, however, extraordinary measures can be taken to keep resources on-line despite operating problems or to return units to service unusually quickly. For example, the availability of Mirant's California generating units was much higher during the western power crisis than had been the availability of these same units under utility ownership,[97] likely a result of the extra-ordinary costs Mirant incurred to keep the units available. A UCAP-type outage performance system with low energy prices during reserve shortage hours would not provide the incentive to incur these extraordinary costs. In addition, availability is not only a function of mechanical forced outages and other kinds of outages may not be independent. These kinds of availability issues are discussed in Section E.

In Fall 2004 ISO-NE proposed a series of revisions to its locational capacity market designed to address this incentive problem by providing incentives for capacity market resources to be available during stressed system conditions. Because the availability incentives under the ISO-NE LICAP design were intended to address a broader set of generator availability issues than simple mechanical forced outages, discussion of this market design is deferred until after the consideration of these broader availability issues in the following section.

E. Availability Limitations

The existing capacity markets in the Northeast focus on transmission system deliverability and operational forced outages and deratings to measure generator availability under stressed system conditions. Experience has shown, however, that generation may be deliverable and in perfect operating condition yet unable to meet load under stressed system conditions because of other availability limitations. There are at least four kinds of problems that can produce this result: fuel availability constraints, energy limits, restrictive start-up conditions, and restrictive availability conditions. The first three of these limitations have figured prominently in reliability crises over the past several years in the Northeast, California and Texas, while the fourth may be of increasing importance as renewable resources are added to the capacity resource mix.

Many of these diverse sorts of non-mechanical limits on generator availability can in principle be taken into account within an UCAP market design simply by requiring

[97] Scott M. Harvey, William W. Hogan and Todd Schatzki, "A Hazard Rate Analysis of Mirant's Generating Plant Outages in California," January 2, 2004.

resources to identify their unavailability due to these types of constraints as forced outages that are reflected in the calculation of the EFORd rating. Thus, it really does not make any difference in the Monte Carlo simulations used to determine the capacity requirement if a resource is unavailable 10 percent of the time due to mechanical forced outages or 5 percent of the time due to mechanical forced outages and 5 percent of the time due to lack of fuel. Moreover, there is nothing inefficient or undesirable about relying on units with low availability rates. From a reliability perspective, twenty 95 MW units with 50 percent availability rates would be preferable to a 1,000 MW unit with a 95 percent availability rate. The probability of losing all 1,000 MW of the single unit is 5 percent, while the probability of having none of the 20 units available is vanishingly small, as long as those outage probabilities are independent.

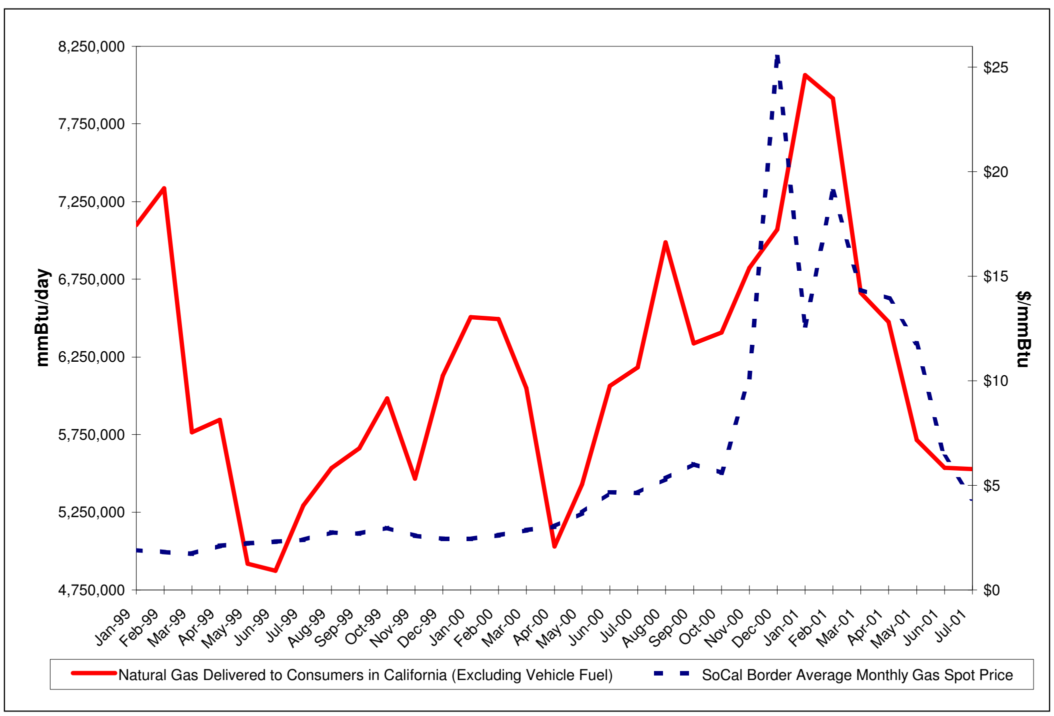
A key characteristic of forced outages in conventional reliability analyses, including Monte Carlo simulations of pool or ISO capacity requirements, is that the outages are treated as independent across units and independent of system conditions. Thus, a pool with 30,000 megawatts of capacity spread across 60 500 MW units with 90 percent EFORd ratings has an expected capacity of 27,000 and would know that the likelihood of having less than 20,000 MW of capacity available would be quite low, because it would be unlikely that 20 of its 60 units would be unavailable during a given day given the overall 90 percent availability factor. The common thread linking the kinds of availability limitations discussed in this section is that they are reasons that the unavailability of capacity resources may be correlated so that the probability of a large amount of capacity not being available on a given day is much higher than assumed in the Monte Carlo simulations used to develop the pool capacity requirement. Thus, if the pool in the illustration above has 10,000 MW with a forced outage rate of 5 percent but this capacity will all be unable to get gas during the 18 coldest days of the year, then the odds that the pool will have less than enough capacity to meet load would actually be extremely high if there was a potential for the pool to have load in the vicinity of 20,000 MW during these 18 cold days.

1. Fuel Availability

While one often thinks of the summer peak as the time of maximum stress on the electric transmission and generation system, several reliability crises have arisen in recent years during the winter months. First, most of the load shedding in California during the 2000-2001 crisis occurred during the winter, not the summer. Second, the last time load shedding was necessary on a wide scale in PJM was during the winter of 1993-1994. Third, during the winter of 2003-2004 NEPOOL came uncomfortably close to requiring load shedding during a winter cold spell. Fourth, one of Ercot's worst recent reliability crises came during the winter of 2003, not during its summer load peak.

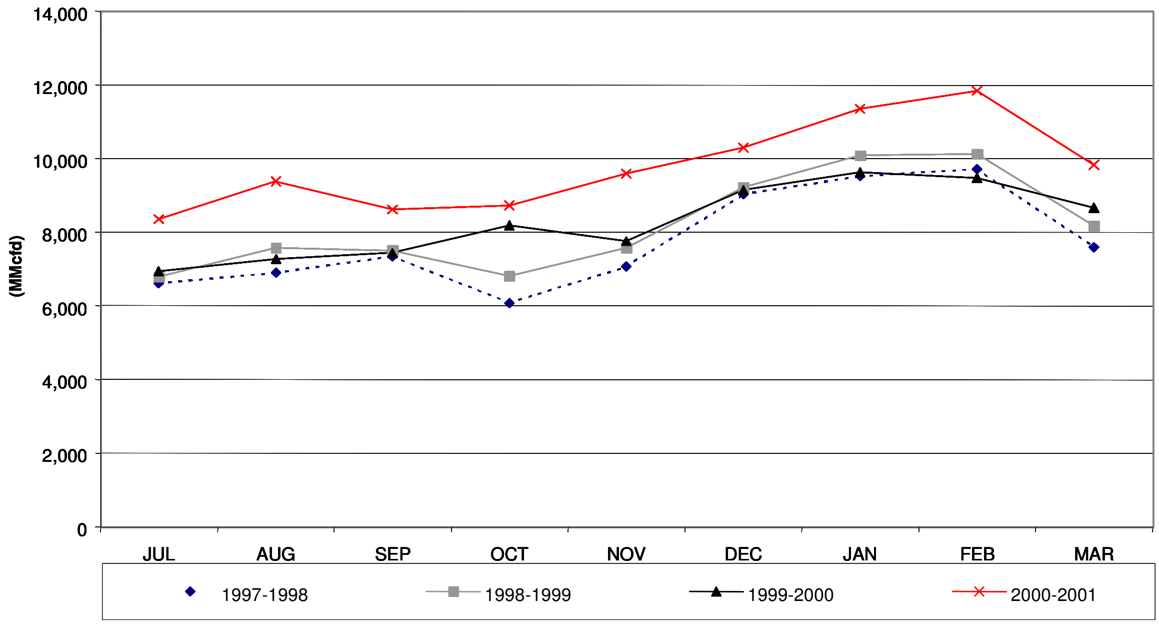
A problem common to all but perhaps the PJM case was that high demand for electricity was accompanied by a high demand for gas for space heating. This high demand for space heating drove gas prices to very high levels, greatly raising the cost of electricity from gas-fired generation and limiting its availability. Thus, in California during the winter of 2000-2001, unusually high gas demand both in California and the west in general driven by low hydro conditions led to very high gas demand and gas prices as shown in Figures 27 and 28.

Figure 27
California Gas Demand and Prices



Source: http://www.EIA.doc.gov/oil_gas/natural_gas/data_publications/naturalgas_monthly/ngm.html and *Gas Daily*.

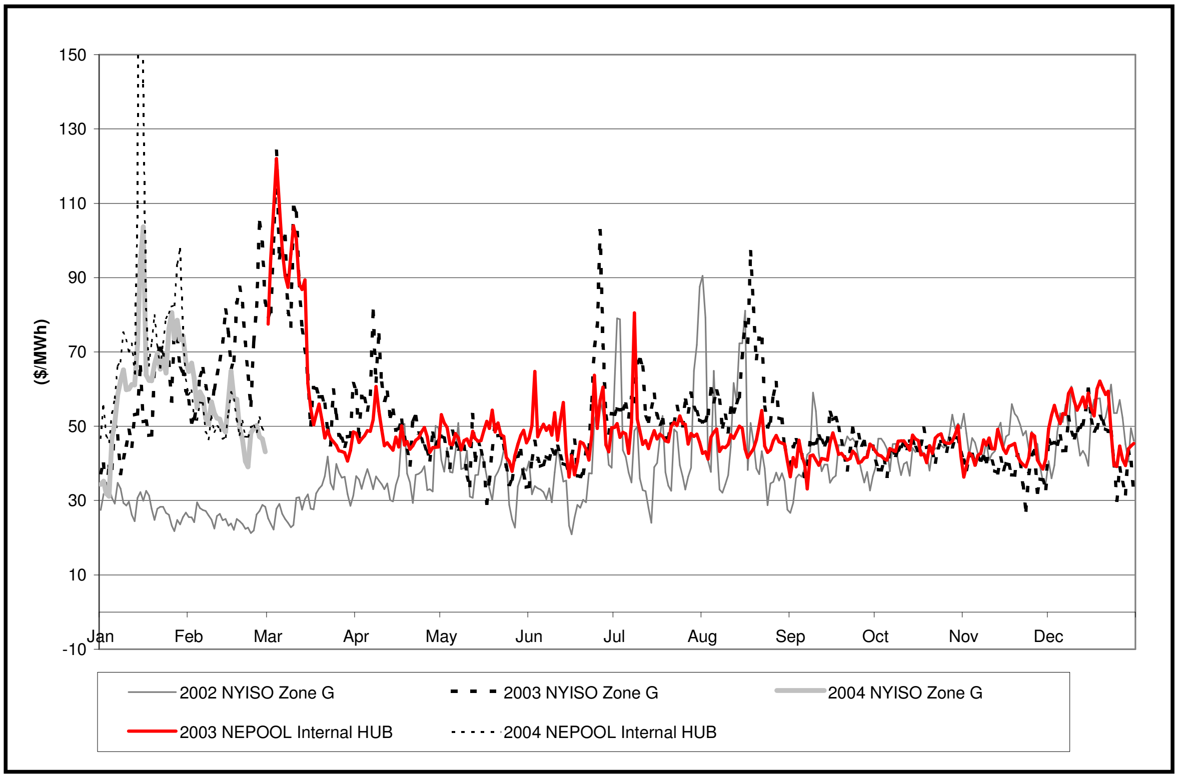
Figure 28
Average Daily Natural Gas Delivered to Consumers in CA, OR, WA, AZ, NV and NM (Excluding Vehicle Fuel)



Source: Table 41 (appended).

Similarly, Figure 29 shows that during 2002 and 2003, high power prices and price spikes have been more common in the New York and particularly New England electricity markets in the winter than in the summer, reflecting high gas prices during severe winter weather.

Figure 29
NYISO Hudson Valley (Zone G) and NEPOOL Hub 2002-2004
Daily Average Day-Ahead market Prices



Source: NYISO.com and ISO-NE.com.

Aside from the impact of high gas prices on power prices, there are three areas of potential reliability impacts of gas shortages under capacity market systems. First, there can be times that gas-fired generation at some locations simply cannot consume any more gas at any price, because higher burns would drop pipeline gas pressure below the critical level leading to generation trips and immediate load shedding.

Second, most existing capacity market systems do not require gas-fired generation to contract for firm gas transmission with either the local distribution company or the interstate pipeline. Under traditional LDC pipeline curtailment rules, a lack of firm gas transmission service meant that a gas-fired generator would not be able to use gas to generate electricity during periods of gas curtailments. In many regions today, however, gas availability is determined by the market, not curtailment rules. In these areas, a generator lacking firm gas transmission service can generate during periods of gas shortage by buying gas at the market-clearing price.[98] The gas system is balanced by customers choosing not to buy gas at high prices rather than by curtailment priorities. Conversely, a generator having firm gas transmission service may sell its gas on days with high gas prices if the electricity price is not sufficiently high to warrant the unit's operation.[99] Overall, physical curtailment is only a concern today in areas with generation served under traditional curtailment rules (at the LDC level). Generators are of course not precluded under a capacity market system from contracting for firm gas transmission service, but a capacity system may diminish their incentive to do so. The crux of a capacity market system is that energy market revenues under shortage conditions are limited by price caps and low costs attached to reserve shortage conditions and marginal capacity is kept available by the capacity payment. If the capacity payment does not depend on having firm gas supply, the incremental energy market revenues may not be sufficient to cover the cost of contracting for firm gas supply and generators may not do so.[100]

Third, gas market price volatility under stressed market conditions may cause gas fired generation lacking dual fuel capability or hedged gas supply to withdraw from the day-ahead electricity market. As noted above, any individual gas fired generator can in principle be assured of obtaining gas under market based gas systems by offering to pay the market clearing price of gas. The generator would then be able to supply electricity at a cost commensurate with the market price of gas. There is nevertheless a reliability problem. In aggregate it is not true that gas fired generators collectively can acquire all the gas they need at the market clearing price. As generators increase their gas consumption at the expense of other consumers,[101] the gas demand of non-generators may at some demand level become highly inelastic in the short run. This may lead to extreme

---

98 Thus, while New England gas LDCs and interstate pipelines curtailed non-firm transmission customers during the January 2004 cold spell, generators lacking firm transmission were still able to obtain gas by purchasing it from entities that had firm transmission. See ISO New England, Inc., Market Monitoring Department, "Interim Report on Electricity Supply Conditions in New England during the January 14-16, 2004 `Cold Snap'," May 10, 2004 (hereafter ISO-NE May 2004). Similarly, although there was generally no non-firm transmission service available to California on El Paso or Transwestern during the winter of 2000-2001, customers lacking firm transmission service could readily acquire daily, bid week, or term gas at the California border (SoCalGas Border pricing point), at the market-clearing price.

99 While ISO-NE found that gas-fired generation with firm gas transmission was somewhat more available than gas-fired generation lacking firm transmission, the difference was not dramatic (56 versus 42 percent) and may have been due to other factors (i.e., gas-fired generators may have firm gas contracts because they are cogeneration units that operate without regard to the electricity market). See ISO-NE May 2004, pp. 68-69, 72, 141.

100 Some reserve requirement systems have required firm gas supply. The MAPP reserve sharing program required either dual fuel capability or firm gas supply and firm gas transportation for capacity to qualify as capacity during the winter season. MAPP Reliability Handbook, Section 3.4.7.2.1.

101 The price responsiveness of non-electric generator gas demand may also vary substantially from LDC to LDC. Some gas LDCs may serve a lot of price-sensitive industrial demand that will reduce consumption as prices rise, while other LDCs may serve largely residential demand that is price-inelastic in the short run.

gas price volatility that could drive unhedged generators out of the gas market and forward capacity contracts would not ensure that enough gas would be available to maintain electric system reliability.

The potential for gas price volatility to reduce generation supply has several elements. First, consider the position of a gas-fired generator operating in a power market that closes after the gas market closes. If the generator purchases gas in the regular day-ahead gas market before offering supply in the day-ahead electric market, the generator risks buying high priced gas that turns out to be uneconomic in the power market. Indeed, this would be likely if gas-fired generators collectively offered whatever it took to buy gas in the day-ahead gas market and then sold their excess gas in the in-day gas market. Alternatively, a generator could offer electricity at a high price in the electric market and then buy gas in-day to cover this position if the generator's offer cleared in the day-ahead electricity market. If gas prices are volatile and the gas market thin and illiquid, however, this strategy could be risky with a substantial possibility of having to pay a much higher than expected gas price in order to cover the electric market position. The same situation could arise if the day-ahead electric market cleared prior to the day-ahead gas market, the generator could sell power before buying gas, but the generator would then risk not being able to purchase gas at a price low enough to make money generating electricity.[102] The best strategy might therefore be to offer power in the day-ahead market at $1,000/MW and to only run to cover this position if is possible to buy gas at a sufficiently low cost. If the gas price were too high, the generator simply would not run.[103]

It may at first appear that these reliability impacts are addressed by the must offer requirement of capacity systems, but this is not the case. A common feature of the NYISO, ISO-NE and PJM capacity markets is that capacity market resources not unavailable as a result of a forced or maintenance outage are obliged to offer their capacity into the day-ahead market of the control area for which they are capacity resources.[104] Gas-fired generators lacking dual fuel capability, lacking firm gas supply, or unwilling to risk purchasing extremely expensive gas are therefore required by the market rules to offer their capacity in the day-ahead market. If these units instead take a forced outage, they suffer a revenue impact in the next capacity market auction. It is

---

102 The market power mitigation system could be another source of risk if it does not track contemporaneous gas prices and generators buying gas on day t to cover generation on either day t or t+1 risk having their offer prices mitigated based on the gas prices reported for day t-1.

103 These kinds of concerns appear to have reduced the supply of gas-fired generation in New England during the January 2004 cold snap. Gas was available for purchase at a price, but the intra-day gas market was thin, and the price volatile and unpredictable. As a result, much gas-fired generation was unavailable due to a "lack of gas supply." See ISO-NE May 2004, pp. 44, 49-51, 56-65, 104-106.

104 NEPOOL Manual for Market Operations p 2-11. PJM Operating Agreement, Section 1.10.1A, Day-Ahead Energy Market Scheduling, Sheet 93; NY ICAP Manual, Section 4.8, p. 4-14; NYISO Services Tariff, Section 5.12.7, Sheets 135c, 135d.

noted above, however, that the financial impact of such outages could be very small for a baseload unit that operates thousands of hours per year and much lower than possible losses from selling uneconomic power in the day-ahead market.

More significantly from a reliability perspective, the ability of units to declare a forced outage when not available due to fuel supply constraints is not consistent with the reliability analysis on which the analysis of control area capacity requirements is based. Control area capacity requirements are based on probabilistic analyses of available generation, transmission and load.[105] Critically, the reliability analyses assume that forced outages are independent events. Because forced outages are modeled as independent events, it is unlikely in these reliability analyses that a large number of generating units will suffer a forced outage on the same day, so many units with low forced outage rates enable a control area to be confident of satisfying its one-day-in-ten-year reliability criteria. If the "forced outage" is actually a failure to offer capacity due to lack of gas supply combined with a lack of dual fuel capability, then the "forced outages" are not appropriately modeled as independent, however; instead they may be highly correlated across many gas fired units lacking dual fuel capability and the control area may have a much higher reliability risk than indicated by the probabilistic analysis used to develop the capacity market requirement. Moreover, as noted above, the actual capacity market revenue impact of a 72-hour forced outage on a baseload gas unit would be very small, providing little incentive for the capacity supplier to incur substantial costs or risks in order to be available.

The mere fact that reliability problems can emerge under extreme gas market conditions is not necessarily a limitation of an capacity market system as these reliability problems could simply be an unavoidable real-world possibility. The problem is that a capacity market system is less effective than forward energy contracts and uncapped day-ahead and real-time prices in providing an incentive for market participants to incur costs to address the potential reliability problems, so a capacity market system may give rise to reliability risks that would not exist under an energy-only resource adequacy system based on effective shortage pricing.

---

105 See Subsection E below.

There are at least three actions that gas-fired generators could potentially take to improve overall reliability that could be impacted by reliance on capacity market versus energy market pricing to assure electric system reliability. First, gas-fired generators could develop and maintain dual-fuel capability. Second, they could inject gas into production area storage, or contract for LNG deliveries into storage making more gas available at times when the pipeline system is constrained. Third, they could contract for new gas transmission capacity into the region, increasing delivery capacity. None of these actions will be incented by a conventional capacity market system.

Often, the best solution to winter gas supply reliability and the market impacts of gas shortages is the development of gas-fired generation with dual fuel capability. At the time of capacity market implementation in the Northeast, a substantial proportion of the former utility gas-fired generation had dual fuel capability and routinely switched to oil during periods of high gas demand. It is not clear, however, whether the current capacity market-based reliability mechanisms in the Northeast will sustain this capability. The capacity market systems currently do not require dual fuel capacity and a considerable proportion of the gas-fired generation that has been built in the Northeast since 1998 lacks dual fuel capability, having neither permits nor oil capable burners. Even in those cases in which generating units were permitted as dual fuel, the generators have not in all cases either installed liquid storage or filled the storage. Worse, from a reliability perspective, there is a prospect for material amounts of the existing dual fuel capable generation being shut down and replaced with gas-only generation having a lower heat rate.

It is noteworthy that California, like the Northeast, used to have substantial fuel switching capability in its electric generating resources. During the 1994 drought in the west, there was substantial fuel switching by California electric generation that did not occur during 2000-2001. Table 30 shows that during the winter of 2000-2001 total electricity generation at the plants in San Diego and Northern California having dual fuel capability was somewhat higher than the total generation at these plants during the same period in 1993-1994.[106] There was a dramatic decrease, however, in the amount of oil-fired generation between 1993-1994 and 2000-2001. There was almost no oil-fired generation during the winter of 2000-2001. This decrease in oil generation increased gas demand from electric generation, raising gas prices.

Table 30
Fuel Shifting in California

	November 1993 – April 1994		**November 2000 – April 2001**	
	Total MWh	**Oil MWh**	**Total MWh**	**Oil MWh**
Northern California				
Potrero 3	529,329	20,580	536,859	0
Hunters Point	629,532	137,329	359,412	0
Pittsburg	4,420,365	251,551	5,402,515	0
Contra Costa	2,111,946	121,611	1,853,595	0
Moss Landing	5,061,748	318,929	3,876,883	0
Morrow Bay	1,774,232	112,484	2,552,311	0
Total	**14,527,152**	**962,484**	**14,581,575**	**0**
Southern California				
Encina	1,261,524	610,662	2,488,493	52,831
Source: EIA Form 759.				

[106] In addition, a number of plants in the LA Basin generated small amounts of power using oil in 1993-1994, and none in 2000-2001.

Part of the change in fuel switching behavior in California between 1994-2001 was due to changing environmental limits and unit capabilities, but another more subtle factor were the changes in gas pricing. In 1994 SoCalGas and PG&E consumers could buy all the gas they wanted at the regulated price and when there was not enough at the regulated price, some customers, including electric generating customers able to fuel switch, were interrupted. As a result, the dual fuel capable electric generators switched to oil fuels every time the gas market got tight in 1993-1994.

By 2000-2001, however, SoCalGas and PG&E, like many other gas distribution companies, had many customers on their system that paid the market clearing price for the gas they consumed. Despite gas demand by electric generation that was far above normal levels due to a combination of reduced hydro generation, nuclear plant outages and cold weather, SoCalGas and PG&E did not have to rely on administrative curtailments to balance supply and demand during the winter of 2000-2001. In 2000-2001, non-core gas consumers in California could generally buy all the gas they wanted at the market clearing price. As long as there was enough gas at the market-clearing price, non-core customers were not interrupted, but the market price rose until non-core gas customers reduced their consumption, making gas available on the spot market. As a result, environmental rules permitting electric generators to fuel switch only when curtailed, rarely came into play because deliveries were rarely subject to curtailment, the gas price rose until the market cleared. The only reason any fuel switching occurred in San Diego during 20002-2001 is that there was no separate locational gas price for San Diego on the SoCalGas system, but San Diego had a separate set of delivery constraints that became binding at times, requiring gas curtailments at the same gas price that cleared the market elsewhere on the SoCalGas system.

Gas-fired generators in the Northeast have generally not encountered these kinds of environmental restrictions on fuel switching but that will potentially end in the near future as some NOx restrictions are extended from the summer to year around.

It is, therefore, important to recognize that with environmental restrictions on fuel switching by gas-fired generation, the gas market price can rise far above the cost of oil before fuel switching occurs. The system may still be reliable in principle if fuel switching can occur if sufficient gas is not available and load shedding would otherwise be required,[107] but market prices can become extremely high under such rules, as was seen in California. Because gas demand may not be highly elastic during winter conditions, the gas price can perhaps become so high that electric generating companies are reluctant to buy gas at those prices out of a fear that they will not be able to recover those costs in the power market.

Beyond the mere possession of dual fuel capability, reliability during winter conditions can also be impacted by the amount of oil fuel in storage. Possession of dual fuel capability by gas-fired generation does not help reliability if the fuel oil stocks are not sufficient to keep the generation burning oil. Problems with oil stockpiles have been an issue in most winter reliability crises. Thus, in Ercot in early 2003, the combination of cold weather and an ice storm in North Texas led to very high gas demand and high gas prices during a period of unusually high winter electricity demand. Many units had dual fuel capability but the severe weather and the uncertainty as to when the cold would abate required husbanding of oil fuel for generation because the impassible roads meant there was no ability for trucks to replenish fuel stocks as they became depleted.

PJM had a similar problem in 1994 when frozen coal piles were accompanied by frozen rivers and ice covered roads that hindered resupply of oil stocks, leading to rolling blackouts by Pepco, PSE&G, Baltimore Gas & Electric, Jersey Central Power, and Vepco. Similarly, ISO-NE reported losing at least 100 MW of oil-fired generation during the January 2004 cold snap due to lack of fuel and additional outages of oil-fired and dual fuel units may have been related to petroleum fuel constraints.[108]

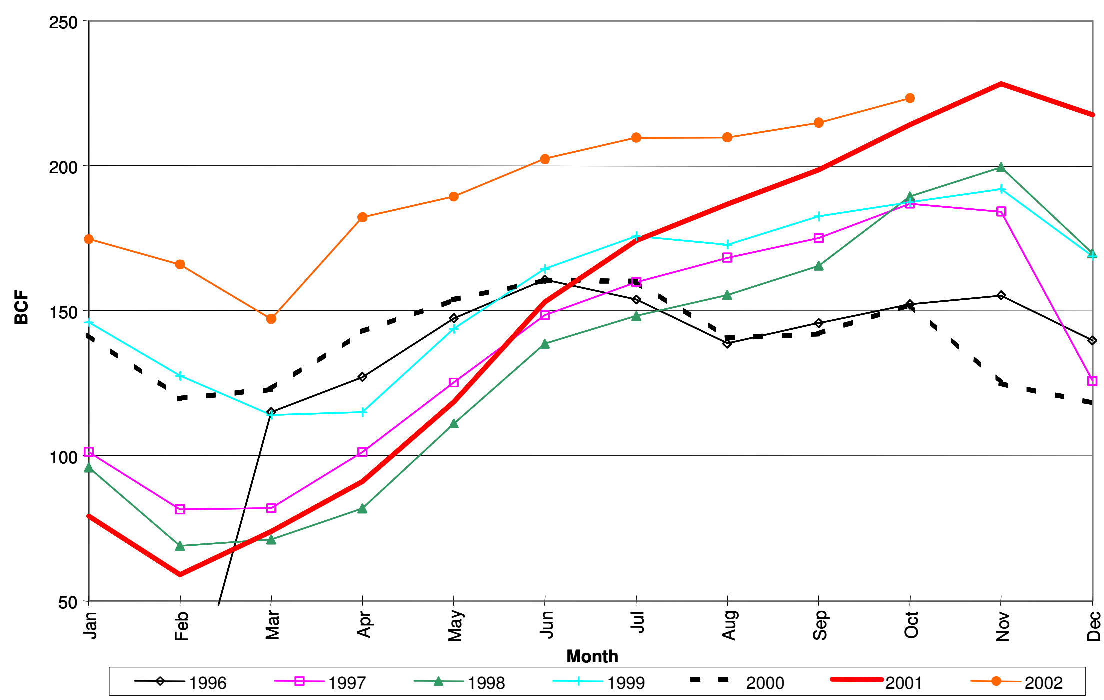
---

107 It is important for the ISO to coordinate operations with the affected gas distribution companies in these circumstances. Unanticipated ramp ups of electric generation can cause reliability problems on the gas distribution system and dual fuel units cannot instantly switch between gas and oil.

108 ISO-NE May 2004, pp. 31, 94-96, 99.

A second potential incentive problem is that capacity market systems such as those in place in the Northeast do not require generators to put gas in consumption area storage to meet generation demand when the gas pipeline system is constrained, but availability of storage gas can be important in meeting load. In Southern California during 2000-2001, for example, the non-core gas customers as a group entered the winter with no gas in storage and therefore no ability to cushion themselves against high gas demand and daily gas balancing rules. This problem would not have been corrected by a conventional capacity market system, but forward contracts for power during the winter months would likely have incented the gas-fired generators to have put gas in storage during the summer to hedge the cost of covering the forward contracts. In New England, LNG terminals provide something analogous to underground gas storage but the capacity market system provides minimal incentive for power generators to contract for LNG capability.

Figure 31
Working Natural Gas in Underground Storage at End of the Month
California (Bcf)



Source: EIA Natural Gas Monthly, Table 14. Before 1998, Table 13.

Finally, a UCAP-based capacity market system does not incent gas fired electric generators to enter into firm contracts for new gas transmission capacity. While power generators helped finance new gas pipeline construction in California in 2001, these contracts were driven by high energy prices and forward contracts, not a capacity market system. A new generator that will obtain most of its revenues from the capacity market or from summer operation, will have little incentive to contract for firm gas transmission capacity to ensure availability of low cost gas in the winter.

There are several ways to address these kinds of fuel supply constraints. One approach would be to add fuel availability requirements to the capacity market program. MAPP has such requirements in its reserve sharing program. The MAPP reserve sharing program requires either dual fuel capability or firm gas supply and firm gas transportation for capacity to qualify as capacity during the winter season.[109] In addition, MAPP requires that units have sufficient fuel storage to enable the unit run during the 4 peak hours five days in succession.[110]

There are, however, a number of complications in applying such an approach. First, since the winter peak is lower than the summer peak in many regions, it is not cost-effective to require that all capacity be available to operate during winter conditions. This could be addressed by establishing a separate winter capacity requirement based on the winter peak and a corresponding winter capacity payment.

Second, in regions with competitive gas markets, it would not be enough to merely require that gas-fired capacity resources contract for firm pipeline capacity as the generators would sell their gas into the spot market if day-ahead and real-time power prices were not high enough to make operation of the resource profitable. Conversely, requiring gas-fired capacity resources to schedule gas day-ahead and then to withhold it from the secondary market would be extremely inefficient on days when the gas is not needed for power generation. Such a rule could lead to extremely high gas prices for gas consumers at the same time that gas pipelines are experiencing operational problems due to large gas injections and small withdrawals by power producers.

A second approach would be to apply some form of derating to resources based on their historic availability during peak conditions. Thus, gas-fired generation that is unavailable during the winter peak might suffer a derating in addition to the random outage derating reflected in the EFORd. In reviewing the January 2004 cold snap, ISO-NE identified the possibility of adjusting its UCAP rating system to more heavily weight outages during peak periods.[111] ISO-NE introduced proposed changes to its locational capacity market system in August 2004 that would more closely tie capacity market

---

109 MAPP Reliability Handbook, Section 3.4.7.2.1. KeySpan has recommended that the Northeast ISOs address this problem by including the cost of firm pipeline capacity or dual fuel capability in the reference price of the installed capacity demand curves. Motion to Intervene and Comments of KeySpan Corporation, FERC Docket No. ER01-3001-014, January 23, 2006. Simply raising the reference price to include these costs, however, would not provide an incentive for capacity suppliers to incur these costs by contracting for firm pipeline capacity or installing dual fuel capability; it would simply somewhat raise the market-clearing installed capacity price in the auction through the operation of the demand curve.

110 MAPP Reliability Handbook, Section 3.4.7.2.1.

111 ISO-NE May 2004, p. 144.

revenues to generator performance during such stressed system conditions. The effectiveness of these proposals in providing performance incentives for capacity market resources is discussed in Subsection 6 below.

A third approach would be adoption of a market design that places greater emphasis on energy market margins to provide generator performance incentives. If electric prices rise during stressed winter conditions to levels reflecting spot gas prices and limited oil fuel stocks, the resulting high power prices would incent gas fired generators to maintain dual fuel capability and incent dual-fired generators to maintain adequate fuel stocks. The potential for high gas prices would incent market participants to fill storage and contract for firm gas transmission. This approach may be difficult to apply, however, in combination with locational market power mitigation utilizing cost-based offer caps if the relevant generators have locational market power. Intra-day gas markets can be thin, and even day-ahead gas markets are thin at some locations, making it difficult to accurately apply offer price mitigation based on spot gas prices during winter conditions when the market price of gas can be very volatile. Moreover, since FERC actions in the California refund case did not allow power generators selling their output in the spot market to retain the benefits from having contracted for firm gas transmission, storage or having purchased gas forward, it is uncertain whether generators will incur such costs in the future unless they are hedging a forward power contract.

2. Energy Limits

The analysis of capacity market requirements and reliability by Eastern ISOs is focused on having enough capacity available to meet demand over short system peaks. Thus, a capacity market reliability analysis normally does not ask whether there is enough energy available to meet load over the year. An important feature of the western electricity crisis over the period 2000-2001, however, was that it evolved from a capacity shortage in the early summer of 2000 into an energy shortage during the winter of 2000-2001.[112] The source of the problem was that the capacity reductions were concentrated among baseload supply, hydro and nuclear generation, which disproportionately reduced the supply of energy. While part of this energy could be made up by other coal and baseload gas units, the magnitude of the energy shortage was such that units with annual operating hour limitations were run until they exhausted their limits and high emission gas fired units were run until NOx allowance requirements effectively precluded further output. Moreover, the need to replace so much baseload power with gas-fired generation contributed substantially to high gas demand that led to transmission constraints on gas deliveries. A different kind of analysis than is typically undertaken in capacity market modeling would be required to assess the risk of energy shortages arising from the

---

[112] The energy shortage was exacerbated by the low level of forward hedging of energy prices, which undermined the solvency of the largest LSEs, further exacerbating the energy shortage when qualifying facilities were not paid for their output and subsequently reduced output or ceased operation.

Western hydro cycle or from multiple nuclear plants outages for a prolonged period within a constrained region or analogous prolonged energy-reducing events (such as a prolonged outage of a major transmission line used to import power).

One way to address energy, rather than capacity adequacy, would be through forward energy contracts that would incent suppliers to take steps to ensure their ability to cover their position following outages. Alternatively, these risks could be analyzed and addressed through a capacity market. The capacity market might, for example, either restrict participation by units with very low levels of hourly energy availability or apply a scale that reduces the capacity payment to generators with annual energy limits below some threshold. The MAPP reserve requirements, for example, have an energy availability provision, requiring that generation other than internal combustion units have permits allowing for at least 500 hours of operation, but this limit is too low to ensure that sufficient output would be available during a sustained energy shortage of the kind to which the West is vulnerable.[113]

One problem with addressing energy adequacy through a capacity market is that reliability does not require that all generation have the ability to run for 6,000 hours per year, it just requires that enough generation be available to avoid running into energy constraints. Reliability requires a load shape of capacity availability; it does not require that all capacity be available for 8,000 hours a year. A capacity market is not well suited to providing these kinds of incentives because capacity requirements apply to all units. Imposing a requirement that all capacity market resources be capable of running for 8,000 hours a year, for example, could greatly magnify capacity costs because much low-cost capacity would be excluded from the capacity market and higher-cost capacity would need to be built and paid for by consumers.

Similar issues can exist for resources with intra-day energy limits such as pumped storage. Both NYISO and ISO-NE currently qualify pumped storage units as capacity providers despite the fact that the units do not produce any net energy. The NYISO and ISO-NE capacity market rules require only that the units be able to operate for at least four consecutive hours each day.[114] While pumped storage units are very valuable in their current proportions in providing reserves and meeting peak load, there is a potential for technology to produce a greatly expanded supply of short duration energy that would qualify for large capacity market payments but have little reliability value at the margin.

---

113 MAPP Reliability Handbook, Section 3.4.7.1.1.

114 NYISO Service Tariff, Section 2.49b, Sheet 36A and 2.74c, Sheet 43; NYISO Installed Capacity Manual, Section 4.8.2; and ISO New England Installed Capacity Manual, Section D1.1.5 and Attachment D, pp. DA-8 and DA-9.

3. Start-Up Conditions

The capacity market systems in the Northeast require that capacity market resources offer their capacity in the day-ahead market. It is not always understood, however, that the generators may accompany those offers with start-up times greater than 24 hours so that the capacity is effectively not available for the next day.[115] This kind of offering behavior can be particularly common for rarely used units that are not expected to operate in the near future and are therefore not manned. In these instances, the long-start up times are submitted to enable the plant operator to recall employees and then start the units. The same kind of behavior prevailed prior to deregulation. Some of the problems during the 1993-1994 PJM cold snap were due to misforecasting of demand and an inability to get units not normally used in the winter manned and on-line. The problem also arose in Texas during the winter of 2003 when some new combined cycles were not available because they had not prepared for such low temperatures, while units normally used for summer peaking had been mothballed for the winter and could not come on-line in time when weather forecasts changed.

Long start-up times can have reliability implications, however, as was seen on May 8 and 9, 2000 when a sudden change in weather forecasts in the northeast caused PJM, New York and New England to be generation short. The NYISO tracked the changes in the weather forecast and by Sunday had a high load forecast for Monday but the NYISO could not commit units with 72-hour start times (on Sunday to be available on Monday), yet these units qualified for full capacity market payments.

Some of these reliability problems arising from demand surprises are unavoidable, but they can be exacerbated by an capacity market system. If most of a high cost rarely operated unit's revenues come from the capacity market and those capacity market revenues will be received even if the unit is unmanned and requires a 72-hour start-up notice, the unit owner has little incentive to staff the unit during normally low load periods and the unit may not be available to meet reliability surprises in the fall, winter or spring. Conversely, if that unit were dependent on high prices in the energy market during shortage conditions for its revenues, or if it had signed forward call contracts that it needed to cover, the unit owner would be more willing to incur higher costs in order to make the unit available to operate on a shorter term basis.

A capacity market system could even potentially sway the choice of generation owners of which units to retire and which to keep in mothballs away from quick start units, if it would be cheaper to keep an old oil or coal-fired steam unit unmanned and in mothballs with a 72-hour start-up time. These kinds of incentive problems can in principle be addressed by modifying the capacity market system so that capacity

---

115 This capacity can be committed by the ISOs but only if they foresee a reliability problem several days in advance.

payments are tied to actual performance during stressed system conditions, as discussed below in Section 5.

4. Limited Availability Resources

A final limitation of UCAP performance systems is their application to resources with limited availability, such as wind and solar generating resources and perhaps also some demand response resources. While energy limited units have the ability, given an appropriate market design, to ensure that their limited energy is used during peak load conditions, wind and solar units are subject to random availability limits. The current UCAP systems are able to accommodate resources with low availability factors, as long as the reduced availability is random, i.e., not correlated with the outages of other resources nor correlated with high load conditions. A mix of intermittent generation resources that were randomly available 10 percent of the time could be counted on for 10 percent of their capacity and would earn proportionately smaller capacity payments under a UCAP system. If the resources had countervailing cost savings, such as the large energy margins the units would earn during periods of high fuel prices, they could be economic in a UCAP system despite receiving a capacity payment based on only 10 percent of their nominal capacity, so diverse types of generating resources can be accommodated within the UCAP capacity market model.

The treatment of wind units under UCAP systems is potentially problematic, however, to the extent that the availability of wind energy is inversely correlated with peak demand, so that reduced availability is not random. This would be the case, for example, if wind energy output at some projects were likely to be lowest on hot humid windless summer days when air conditioner load is at its maximum. Solar energy output is likely to be positively correlated with peak load conditions in the summer but output may be inversely related to winter peak conditions.

While the non-availability of wind energy or solar power in the winter can be handled as forced outages under UCAP systems, this is not sufficient for the purpose of reliability analysis. As noted above with regard to fuel availability, forced outages are treated as random events in the reliability analysis used to develop UCAP requirements, but this treatment will not be accurate if wind energy non-availability during stressed system conditions is not random but correlated with high demand conditions and correlated across wind units.

NYISO and ISO-NE calculate UCAP for intermittent resources based on an historical capacity factor adjusted to remove the effects of outages.[116] PJM calculates the UCAP for intermittent resources based on a historical capacity factor during summer

---

116 New England Power Pool, Market Rule 1, Section 8.3.6. NYISO Service Tariff Section 5.12.11(d), Sheets 142, 135B, 15B.01. NYISO is evaluating its methods for evaluating the availability of wind resources.

peak hours (HE 15, 16, 17 and 18).[117] The PJM approach can to a degree capture the correlation between low output and hot, still summer conditions, but some of the hours included in the calculation of the capacity factor may be low load, windy conditions. It may be necessary to restrict the calculation of the capacity factor to the high load summer peak hours, rather than all peak hours, to accurately measure the contribution of wind units to meeting summer peaks. Once again, it is potentially difficult to account for the reliability impact of non-random outages in an UCAP system.

5. Discussion

The common feature of all of these considerations is that the current UCAP systems do not provide enough incentive for capacity market resources to be available during stressed system conditions and availability is not exogenously determined by random forced outages but depends on the choices made by the resource owner in incurring costs in order to have the resource available. This is intrinsic to the current UCAP-based capacity market systems. These systems necessarily pay the generator less for being available during the shortage hours than the generator would receive during shortage conditions under an energy only market with appropriate shortage pricing. UCAP systems attempt to compensate for this incentive problem by requiring that resources receiving compensation for providing capacity demonstrate their generating capacity. This approach works well for many thermal resources as if the capacity exists, forced outages will be random (and this random risk can be analyzed) and the resources not unavailable due to forced outages do not need extra incentives to operate during these shortage conditions. This approach is not adequate, however, to deal with fuel availability limits and start-up conditions in particular, as resource availability under strained system conditions depends on choices made by the resource owner, and resource owners will not incur the efficient level of costs to maintain availability if they realize small returns from incurring those costs.

On the other hand, energy market prices (for both energy and reserves) can provide appropriate incentives for dual fuel operation, keeping oil in storage, etc. It is also possible that some of the problems experienced in Texas and NEPOOL may be transitional issues with new generation and new market conditions. Even absent modification of the NEPOOL capacity market system, we might see developers focusing on providing dual fuel capability. These complications in using an capacity market system to ensure that both energy and capacity are available during stressed system conditions are leading to further evolution of Eastern capacity markets, but the evolution is proceeding in very different directions in PJM (RPM), ISO-NE (locational capacity markets with stronger availability penalties) and NYISO (greater reliance on shortage

[117] PJM Manual 21, Rules and Procedures for Determination of Generating Capability, Appendix B, April 30, 2004.

pricing in the energy market) and it is not clear which if any of these approaches will prove to be workable and successful.

6. Performance Incentives for Capacity Resources

Over the 2004-2006 period, ISO-NE developed and refined a series of proposals to address the kind of availability problems associated with both forced outages and fuel availability, energy limits, start-up times, and restrictive availability conditions by modifying the UCAP performance standards so that the receipt of capacity payments would be tied to performance during stressed system conditions. These proposals evolved over time and have yet to be developed into complete tariff language so the discussion below does not describe every feature of the various proposals nor does it identify every potential problem with every variation in the proposals but instead focuses on discussing a few key features of these proposals; issues that are important in developing improved availability incentives within a capacity market system for maintaining resource adequacy.

Critical Hours

The initial fall 2004 formulation of the ISO-NE LICAP availability metric provided that capacity payments would be reduced based on whether a capacity market resource was actually on line, generating energy in real time (or available to generate within 30 minutes) during roughly 100 critical hours. In effect, therefore generating units would receive their capacity payment for being available during these 100 critical hours.[118] Generators that were unavailable during these 100 hours, whether as a result of a mechanical forced outage or because they lacked fuel, could not be started up in time or for whatever other reason, would not receive the capacity market payment. Since capacity resources that were not available to provide energy or reserves during these critical 100 hours would lose the capacity market payments, capacity resources would have an incentive to incur costs in order to assure availability during these critical hours whether the hours were on hot summer days or during winter gas prices spikes. Not all units might find it cost effective to incur the cost of being available during all of these hours but there would be a balance between the cost of making additional capacity available during a period and the probability of some of these critical hours falling during that period.

With performance incentives based on availability during critical hours, resource owners would also need to balance cost savings from setting long start-up times for unmanned units with the potential forgone capacity market revenues if the unit were as a result unavailable during these critical hours. Resource owners might still leave units

---

118 Stoft 2004, pp. 96-101.

unmanned with 72 hour start-up conditions during shoulder months but they would have an incentive to take steps to carefully monitor weather and outage conditions and man the units to bring them on-line as soon as possible if conditions changed.

Such a performance based availability incentives could, in principle, also provide consistent incentives for resources with restrictive availability conditions, as such units would receive capacity market payments based on their actual availability during the critical hours.

While such a mechanism would provide incentives for avoiding forced outages during peak load conditions similar to an energy-only market it would also have some undesirable features. In particular, as originally formulated, the 100 critical hours during the year used to determine performance would not all have been hours of reserve shortage. The original proposal would have defined the 100 critical hours to have been the hours of reserve shortage plus additional hours selected based on the ratio of the real-time energy price to the "daily spot price" (presumably meaning the day-ahead gas price).[119] It is the authors understanding that for capacity planning standards based on the traditional 1 day in 10 year load shedding criteria, the expected number of annual hours of reserve shortage lies in the range of 5-20 hours. The vast bulk of these 100 critical hours would therefore have been hours with higher than average marginal heat rates but no reserve shortage. Some of the critical hours would have been high load summer days on which there was no reserve shortage but high heat rate units had to be used to meet load. Other critical hours would be on winter days on which cold weather drove the intra-day gas price well above the day-ahead gas price and similarly elevated real-time power prices relative to the day-ahead gas price.

The application of such an availability metric could have lead to significant operating inefficiency during critical hours with high marginal heat rates but no reserve shortage. If a $50,000/year capacity payment were contingent on a unit generating during the 100 critical hours of the year with the highest ratio of the real-time price to the day-ahead gas price, these payments would amount to $500/MWh subsidy for generating output during these hours. Not only would this incentive be inefficient but since by definition not all available generators could be generating during an hour in which there was more than enough capacity available to meet load, such a rule could materially distort prices.

The potential distortion in energy prices from such a performance incentive system for capacity resources could be reduced by making the capacity payment contingent on the resource either generating energy or providing reserves. There would still be a potential distortion in energy and reserve prices (since again by definition not all capacity is needed to either generate energy or provide reserves during non-shortage

---

[119] Stoft 2004, pp. 98-99.

hours).[120] The efforts of capacity market suppliers to force their units into the dispatch during likely critical hours in which there was no reserve shortage would not only reduce the efficiency of the economic dispatch but would also depress real-time prices making it less likely that those hours would be designated a critical hour. The inefficiency introduced by the competition to operate during non-shortage critical hours would have been reflected in capacity offer prices and the cost of the dead weight loss would have ultimately been borne by consumers.

Another alternative in implementing performance incentives for capacity resources would be to count capacity as available if it is offered for dispatch in real-time, even if it is not actually used to either provide reserves or generate energy. This approach would eliminate the incentive for generators to force themselves into the real-time dispatch by bidding below cost but would still provide an incentive for capacity market resources other than quickstart units to be on line at minimum load during periods of high gas prices to assure receiving the capacity payment. This incentive would not only be inefficient, it could undermine gas system reliability by providing an incentive for units with high minimum load heat rates to be on line burning gas at times of stress on the gas pipeline system.

Shortage Hours

The ISO-NE LICAP availability metric evolved to address these potential distortions by limiting the application of the metric to hours of actual real-time reserve shortage.[121] Since the system operator by definition needs all available and dispatchable capacity to be on-line (or providing reserves available within 30 minutes) during hours of reserve shortage, all available and dispatchable capacity would either be dispatched to generate energy or scheduled to provide reserves and the kinds of distorted operating incentives associated with the ISO-NE LICAP proposal would be avoided under such an availability metric.

The general concept of providing capacity market suppliers with incentives tied to their actual availability during reserve shortage conditions is a desirable evolution of capacity market based resource adequacy mechanisms, but there are a number of complications in implementing such a system of performance incentives. ISO-NE's FCM proposal is used to illustrate these issues but the discussion is focused on general issues rather than the details of the FCM proposal, many of which have not yet been

---

120 Similar observations have been made by many others in proceedings related to the ISO-NE LICAP proposals; see, for example, Cliff W. Hamal, Answering Testimony, FERC Docket No. ER-563-030, November 4, 2004, pp.13, 91-93.

121 ISO-NE Initial Brief, pp. 49-50. FCM Proposal, Explanatory Statement, Section III.A, Settlement Agreement Section VC2. David LaPlante, Rebuttal Testimony, Docket ER03-563-030 (hereafter LaPlante Rebuttal), pp. 14, 86.

worked out. The first issue is the related choices of whether performance incentives should be a prospective or retrospective system and whether performance should be measured on an incident by incident basis or on a proportionate basis. A second and related issue is the measure of capacity used to determine capacity payments by consumers. The third issue concerns the differential impacts on the performance of day-ahead markets and the unit commitment process of energy-only and capacity-market-based resource adequacy mechanisms. A fourth issue concerns the treatment of planned maintenance outages in the performance system. The fifth issue concerns the potential relationship between the deliverability requirement and the performance incentives. A sixth issue concerns the treatment of energy limited units. Capacity-market-based resource adequacy mechanisms potentially give rise to problems in defining the obligations of limited energy resources that raise the overall cost of meeting load by raising the risks associated with meeting load with energy limited units. The seventh issue concerns the relative performance of capacity market based resource adequacy mechanisms and energy-only (actually energy and reserve) pricing system during periods of energy, rather than reserve shortages.

Prospective or Retrospective

An initial issue in defining capacity resource performance incentives is whether the performance incentives should be based on a prospective or retrospective system and whether performance should be measured on an incident by incident basis or on a proportionate availability basis. The first element of these choices is whether poor performance during the current capacity market period will impact current capacity market revenues or will affect capacity market revenues in the next period.

ISO-NE's FCM system would take this approach, adjusting capacity payments during the current year based on the availability performance of the resource during the same year. One approach would be to base the performance incentive on the expected number of reserve shortage hours in the capacity resource simulation. A supplier would forgo a per-megawatt penalty for each shortage hour that it was unavailable, regardless of whether there were 2 or 40 such hours during the year. Obviously, if the capacity payment has a lower bound of zero, there would be no incremental performance incentive under such a system once such a capacity resource unit had been unavailable for the shortage hours required to reduce the capacity payment to zero.

Specifically, the ISO-NE FCM proposal would reduce the capacity payment to a capacity resource based on the proportion of its capacity that is unavailable during each reserve shortage event during the current capacity year. Specifically, under the current version of the FCM proposal a resource would lose 5 percent of the annual capacity payment (FCA) for failure to be available during a given reserves shortage event. This penalty would will increase by 1 percent per hour for events that exceed 5 hours up to a maximum per day penalty of 10 percent of the annual capacity payment. In addition, the

FCM proposal provides for a maximum penalty in a month equal to 2.5 times the monthly capacity payment (or roughly 20 percent of the annual capacity payment).[122] Total performance penalties would be capped at the capacity payment, less other adjustments.[123]

One complication in applying such an ongoing reserve shortage hour performance metric is that the number of hours of reserve shortage can vary widely from year to year and the actual number of reserve shortage hours will not be known until the end of the year.[124] If the penalties under an FCM type incentive system were set such that a resource that was unavailable during every reserve shortage hour would earn no capacity payment in a year having the expected number of reserves shortage hours then the performance incentives would be appropriate for many units. For example, if there are an average of 20 reserve shortage events per year, then a resource than was available during two of them would suffer two 5 percent penalties and earn 90 percent of the capacity payment. If there were only 10 reserve shortage events during 25 percent of the years and 30 reserve shortage events during another 25 percent of the years, the unit would earn 95 percent of the capacity payment in some years and 85 percent in others, but it would average out to earn 90 percent of the capacity payment.

A significant caveat is that this equivalence will only hold for resources with sufficiently high availability levels that the capacity payment is never truncated at zero. Suppose, for example, that a resource had a 20 percent availability factor. Then in the example above the unit would be expected to be unavailable during 8 of the reserve shortage events in the years with 10 reserve shortage events, losing 40 percent of the annual capacity payment. In the years with 30 reserves shortage hours this resource would be expected to be unavailable during 24, which would result in penalties of 120 percent of the annual capacity payment. If the penalty is capped at 100 percent of the capacity payment, this resource would average earning 30 percent of the capacity payment over the two years, while only being available during 20 percent of the reserve shortage events.[125]

---

122 See FCM Settlement Agreement, Item 11 Section VC2. It is not clear what is actually intended as Item 11. Section V.B.2 provides that the monthly capacity payment less peak energy rental adjustment cannot be negative, which is inconsistent with the maximum penalty being 2.5 times the capacity payment. There are a variety of additional rules to handle special situations such as a rule that the maximum penalty from a single outage spanning 4 days or less but spanning 2 months is 2.5* the monthly payment, just as if the outage fell within a single month.

123 In particular, the sum of the peak energy rental deduction and the outage deduction is capped at the total capacity payment so that the capacity payment cannot become negative.

124 These comments focus on the basic concepts embodied in the FCM proposal; they do not describe or evaluate all of the detailed provisions or ambiguities.

125 In the extreme case, a resource that was never available would forgo 50 percent of the capacity payment in the year with 10 shortage hours and all of the capacity payment in the year with 30 shortage hours, earning 25 percent of the payment on average without ever performing.

Perhaps because of this asymmetry for low availability resources, the ISO-NE FCM proposal has special rules for poorly performing units. These rules provide that resources whose availability has been less than 40 percent during reserves shortage events in three years over a four year period will lose their ability to participate in future capacity auctions.[126] A limitation of this approach to applying a performance standard with such "poorly performing units" rules is that there actually nothing undesirable about such "poorly performing units." "Poorly performing resources" may be quite valuable if the capacity can be supplied at low cost and the resource's outages are independent of the outages of other resources and of shortage conditions. The problem is not with the performance of the resources but with the performance of the incentive system. Because of the limitations of the FCM incentive system when applied to low availability resources, the FCM incentive system would not work as intended if applied to intermittent resources. The ISO-NE FCM settlement therefore has provisions that exempt intermittent resources such as run of river hydro and wind from both the performance penalties and the poorly performing unit rule, providing that another set of rules and performance incentives will be developed to apply to such units.[127]

The FCM proposal does not include a mechanism for applying some form of performance incentive to intermittent resources and demand response, providing that these rules will be developed in the future.[128] This will not be straight-forward within the FCM framework because of the large capacity payment relative to the expected value of capacity and the potential for resources with low expected availability to earn large capacity payments in years with few reserve shortage events. One way to address these kinds of issues would be to allow capacity payments to go negative, which would be analogous to the outcome in an energy only market for a resource that entered into a forward contract, but intermittent resources might not enter into forward contracts in an energy only market, they might simply sell in the spot market, earning high margins when they operated during reserve shortage conditions, but not risking large losses if they happened not to be generating during a series of high priced reserve shortage hours.

Another way to apply such a on-going performance standards that would be applicable to low availability resources would be to base the capacity payment for a resource on the proportion of the year's reserve shortage hours or events for which the resource was available. Thus, a resource that was available during 10 percent of the reserve shortage hours would receive 10 percent of the capacity payment and a resource that was available for 90 percent of the reserves shortage hours would get 90 percent of the capacity payment. Ten 100 MW resources that were available 10 percent of the time would be paid for 100 MW of capacity, as would a 125 MW unit that was available 80

---

126 Settlement Agreement, Item 11, Section VC7. There are also rules for poorly performing units to re-establish their right to sell capacity.

127 FCM Explanatory Statement Section III A p. 12; Settlement Agreement, Item 11, Section V.C.5.

128 Settlement Agreement Item 11, Sections II.E, V.C.8.

percent of the time. A limitation of this approach is that the performance incentive would vary with the actual number of shortage hours, the availability incentive from a $50,000/MW year capacity payment could range from perhaps $25,000/hour in a year with only two shortage hours, to only $1,250 per hour in a year with 40 such hours and perhaps only $250 per hour in a year with 200 shortage hours.

This approach to performance standards might be appropriate for resources whose availability was largely outside the control of the revenue owner. The per-incident cost of not being available might vary from year to year but if availability is exogenous, depending on wind, sun or run-of-river hydro conditions, these imperfect incentives might not be a serious limitation. This capacity market approach to performance incentives requires a centralized approach to defining performance for all such resources. Under an energy-only resource adequacy mechanism, the resource owner could determine the extent to which it sold forward. For example, in an energy-only market design, some intermittent resources might find it profitable to combine their resource with some sort of energy storage system so as to firm up their supply and enter into forward contracts.

A particular problem with such a performance incentive system is that the variation in incentives would be prospective, not only retrospective. Thus, if there had been 40 reserve shortage events during the summer, resource owners would know that performance during additional reserve shortage events in the fall would have relatively little impact on their capacity payment so they would have less incentive to incur significant costs to be available during an additional fall reserve shortage event. Conversely, during a year with a cool mild summer there might not have been any reserves shortages, so capacity suppliers would have an incentive to incur extraordinary costs to be available during possible reserves shortage hours during the remainder of the year as their entire capacity payment could potentially hang on their performance during a single reserve shortage event.

These limitations of ongoing performance incentives in a capacity market system would not present in an energy only market, because each resource would generate shortage revenues during each shortage hour in which it was available, regardless of whether it was on average a low or high availability resource and regardless of how many other shortage hours there had been during the year. If a resource signed a forward contract and failed to be available during a large number of shortage hours during the year, such a low availability resource could generate negative margins. Incentives similar to those existing in an energy-only market with shortage pricing for suppliers entering into forward contracts could be provided for under a capacity market, including appropriate incentives for low availability units, by permitting capacity market payments to be negative for resources with a large number of outages during a year with many reserve shortage hours.[129]

Another approach to providing performance incentives within a capacity market system would be to average availability performance over a period of time and make capacity payments based on historic average availability during a given number of past reserve shortage hours. Under such a prospective performance incentive system, a resource that had been available during 90 percent of the last 50 shortage hours, for example, would receive 90 percent of the capacity payment for the next period.

The UCAP system is an example of a prospective system, as poor outage performance in the current period affects future rather than current capacity market payments, with the amount of lag varying across the various ISO capacity market designs. One advantage of such an approach is that there would be no discontinuity in the capacity payment or ex ante incentives between years with an unusually low or high number of hours of reserve shortage. Such a system could also accommodate "poorly performing resources" including intermittent resources. A prospective capacity payment system based on reserve shortage hours would have the potentially unattractive feature, however, that since the performance factor would be adjusted only when reserve shortages occurred, capacity resources could have their performance factor, and thus their capacity payment, frozen for long periods of time if a summer with many hours of reserve shortage were followed by a couple of mild summers with few reserve shortage hours.

The difference in incentives between a current and prospective performance incentive system could be particularly important for limited energy resources. During a year with many hours of reserve shortage, such as a low hydro year in the west, part way through the year limited energy units might have incurred performance penalties calculated relative to the expected number of reserve shortage hours that would eliminate their entire annual capacity payment. Under an ongoing performance system, such units would have no incentive to incur costs in order to restore their availability until the end of the capacity market period unless the capacity payment could go negative. For example, during the western energy crisis the owners of a number of limited energy units in California worked with regulators to modify the permit conditions to enable units that had exhausted their annual operating hour limit to return to operation, typically with fairly high costs for the extra hours. There would be no incentive under an ongoing system of availability incentives for a resource owner to incur such costs once the resource had incurred performance penalties that eliminated the annual capacity payment.

---

[129] As observed above, however, resource owners in an energy-only market could choose whether to sign forward contracts and would receive high spot market prices during shortage conditions whether or not they signed a forward contract. Under a capacity market system with performance penalties that can go negative, suppliers in effect must sign a forward contract in order to be paid the market price of power during shortage conditions.

The underlying difficulty with all of these approaches to providing efficient performance incentives in a capacity market system is a fundamental feature of capacity market resource adequacy systems, the timing of the payment for the capacity is different from the timing of the reserve shortage events. Thus, capacity resources will earn the same capacity payment in years with many reserve shortages and in years with few reserve shortages. It should be kept in mind, however, that a forward energy contract in an energy only resource adequacy system would be in a similar position. The seller would receive the same energy payment in years with many reserve shortages as in years with few reserve shortages. If we think of the difference between the energy contract price and the seller's incremental generating costs as the capacity payment, the seller under a forward energy contract may earn basically the same implicit capacity payment during years with few reserve shortage hours as during years with many reserve shortage hours. The seller has more outage risk during the year with many reserves shortages, however, and under such a forward energy contract in an energy only market with reserve shortage pricing, the seller could lose more than its implicit capacity payment were it to suffer a sustained unit outage during a summer with many reserve shortage hours. Eliminating the lower bound of zero on the capacity payment would permit capacity market performance incentives more closely aligned to those under energy only pricing with forward contracts and would eliminate the problems in accommodating "poorly performing units" and intermittent resources.[130]

Defining Capacity

A second issue in applying performance incentives concerns the way the capacity is measured for the purposed of defining the capacity requirement (and applying the capacity market demand curve under demand curve systems). Under the NYISO UCAP model, the capacity requirement is defined in terms of nominal capacity reduced proportionately to reflect the historical forced outage rate. In effect the UCAP quantity purchased is the amount of capacity that is available on an expected basis. The capacity of a resource that is only available 50 percent of the time, would therefore be discounted 50 percent in calculating the amount of capacity purchased, whether the capacity demand curve were vertical or sloping. As observed above, the total capacity payments by consumers are fixed under a UCAP system, equal to the capacity payment times the amount of UCAP purchased.

Although the discussions of the LICAP demand curve and the FCM capacity target (ICR) are not clear on this point, it appears that the determination of the amount of capacity required to maintain reliability would be based on an assessment of the

---

130 Applying performance incentives that allow the capacity payment to go negative could, however, exacerbate the adverse impacts of undefined performance obligations under capacity market systems, particularly for energy-limited resources as discussed below.

availability of the resources used in the Monte Carlo simulations.[131] An important difference from the UCAP capacity market design, however, is that the under the ISO-NE LICAP and FCM designs, the capacity payment would apparently be based on the nominal supply of capacity, adjusted for outages rather than on the expected supply of capacity. The FCM settlement proposal has language suggesting an intent to adjust the actual quantity of capacity targeted for purchase in the auction based on historical outage experience.[132] The issue considered here is different; it is how the quantity of capacity on which the payment is based is defined, is it expected capacity or nominal capacity. Thus, under a UCAP system a 100 MW resource with a 90 percent availability factor would be able to supply 90 MW of UCAP and would be paid the market clearing price for this UCAP, with the offers in the UCAP market reflecting that fact that 100 MW of physical capacity would need to be maintained in order to be paid for 90 MW of UCAP. Under the FCM design, the 100 MW resource would be entitled to a capacity payment for all 100 MW of nominal capacity,[133] but the payment would be subject to reduction based on the actual availability of the resource. Thus, if the resource were available during 90 percent of the reserve shortage hours it would collect 90 percent of the capacity payment for 100 MW of nominal capacity. From this perspective the UCAP and FCM designs would operate similarly. There are, however, a couple of important differences.

First, unlike the fixed year to year UCAP payment, if capacity payments were based on nominal capacity adjusted for performance with per-outage penalties, the payments by loads would potentially vary from year to year, as there would be few availability penalties in a year with few hours of reserve shortage. Thus, resources with an average 90 percent availability factor might be paid 95 percent of the capacity price in years will few shortage hours and 85 percent of the capacity price in years with many shortage hours. This variability of the capacity price from year to year might be somewhat unattractive to loads, particularly the higher payments in the years with few shortage hours. If there were a significant quantity of low availability factor resources in the market, the total nominal MW of capacity resources could be materially larger than the expected quantity of capacity. There would be a potential for consumers to pay materially higher than normal capacity charges in years with few or no shortage hours and thus no availability penalties.

---

131 David LaPlante, Prepared Direct Testimony, FERC Docket No. ER03-563-030, August 31, 2000, pp. 27-28, 39-42; Stoft 2004; Settlement Agreement, Item 11, Section III.B. Thus, the LICAP and FCM models contain the same non-sequitur as the existing UCAP models; the capacity requirement for each region is calculated based on implicit assumptions regarding which resources (with what level of reliability) will provide capacity within each region.

132 Settlement Agreement, Item 11, Section III.B.1.

133 Settlement Agreement, Item 11, Section III.B, “The ICR (installed capacity requirement) purchased shall be based on the summer seasonal claimed capability.”

Second, the FCM settlement agreement avoids this variability of consumer payments through a provision that rather than the outage payments by generation resources serving to reduce capacity payments by consumers, consumers would apparently pay the full capacity payment regardless of the resource outage rates with the performance penalties of capacity resources flowing to those capacity resources that were on line during reserve shortage events.[134] This appears to be a very important provision of the FCM proposal for a couple of reasons. First, with such a rule, total capacity payments by consumers are fixed from year to year and do not vary with generator outage rates. Second, however, this provision means that total payments by consumers are determined by the total nominal capacity of capacity resources, rather than the expected capacity of capacity resources. Thus, if the resource mix shifts to resources with lower availability factors, the amount of capacity paid for by consumers would increase, because the total amount of capacity purchased would rise.

If the capacity market is competitive, these capacity market revenues in addition to the basic capacity payment would be taken into account by capacity suppliers in their auction offer prices.[135] Suppliers would, in consequence, offer capacity at lower nominal prices than would be the case absent this provision. Nevertheless, this design will have some impacts on market outcomes. This feature of the FCM proposal means that either an existing resource in determining its delisting price or a new resource in determining its offer price, needs to have an assessment of the expected level of these capacity penalty payments. While a capacity supplier may have a good assessment of the reliability of its own resource, the magnitude of these penalty payments requires will depend mostly on the reliability of the resources supplied by others. Given the variability in outages and in the number of reserve shortage hours it may be difficult for resource owners to assess the expected level of these penalty charges for the next year. This assessment would be even more difficult in the context of the FCM auctions which would be occurring several years in advance of operation, requiring that bidders take into account the changes in the mix of generation resources and reliability over the next several years in estimating the level of future penalty revenues. By thus complicating the bidding process, this kind of a capacity market feature may have a variety of effects on the capacity market whose impact is difficult to assess.

The bidding complexity introduced by this feature of the FCM market design would impact the role of the market monitor as the market monitor would need to formulate an assessment of the expected level of these penalty charges and revenues in evaluating the competitiveness of supply offers from both existing units and new capacity. All of these assessments will be much more difficult and uncertain than the

---

134 Settlement Agreement, Item 11, Section V C 3.

135 Or in their willingness to remain in operation if offer prices do not set capacity prices (as discussed in Section III.H).

evaluations required by suppliers entering into forward contracts within an energy-only resource adequacy design.

A third impact of this approach of defining the target in terms of nominal capacity is that there will be a potential mismatch between the amount of capacity required to maintain reliability and the amount of capacity purchased in the three-year out auction. If the quantity of capacity purchased in the three year out auction is defined in terms of nominal summer capacity, the amount of expected capacity available during reserve shortage conditions will depend on which resources are purchased in the auction. The ISO's capacity target from a reliability standpoint would be formulated in terms of the expected amount of capacity available during stressed system conditions, which would need to be translated into a nominal capacity based on an expected availability rate. The expected availability for the capacity actually purchased in the auction would depend on which capacity is purchased and would be higher or lower than assumed.

Impact on Day-Ahead Markets

Now, consider the third issue, the differential performance of day-ahead markets under energy only and capacity market based resource adequacy systems. While a system of capacity payments tied to availability during hours of reserve shortage would provide clear benefits relative to the incentives under the current UCAP availability systems, such a capacity system would provide incentives that are markedly different in some respects from those of an energy-only market with shortage pricing. Suppliers whose capacity was not on-line during real-time reserve shortage hours because the reserve shortage was not anticipated and the supplier's unit was not committed by the system operator in the day-ahead market would forgo the capacity payment for these hours of reserve shortage just as if the resource had sustained a forced outage. At one level this is analogous to the incentives under an energy only market as resources that were not on line during an hour of real-time reserve shortage because they had not been committed in the day-ahead market would similarly forgo recovering their costs in these shortage hours. The potential for real-time surprises leading to reserve shortages would provide an incentive for suppliers to invest in quick start capacity in either an energy-only market with shortage pricing or a capacity market with performance incentives based on availability during real-time reserve shortages. There are, however, some important differences that need to be kept in mind in comparing availability incentives in energy-only and capacity markets.

If real-time energy prices during shortage hours can reach $8,000/MWh as opposed to only $1,000/MWh, this difference in the level of real-time prices will affect the bidding by consumers and virtual traders in the day-ahead market. Capping real-time prices at $1,000 reduces the costs to a consumer of underbidding its load in the day-ahead market and would similarly reduce the financial returns to virtual bidders taking long positions in the day-ahead market. The impact on day-ahead market incentives can be particularly important on days when the load forecast is uncertain, as the lower the financial cost of underestimating real-time load, the less incentive for LSEs to bid more than their expected load into the market to protect against potential shortage conditions.

The day-ahead financial market therefore has a potential to systematical commit less capacity during uncertain days under a capacity market system with outage incentives than would be the case under an energy only market with shortage pricing in real time. A capacity market system with reserve shortage based performance incentives such as the ISO-NE LICAP or FCM proposals would provide capacity suppliers with an incentive to try to ensure that their units were on-line during days with uncertain demand forecasts and perhaps with underscheduling by loads, but it should be anticipated that suppliers will not have as good information about likely power consumption as would the actual load serving entity. Indeed, much detailed load data is not publicly available and therefore could not be used by suppliers in forecasting and the potential for real-time reserve shortages. The original ISO-NE LICAP proposal in effect placed the responsibility for avoiding real-time reserve shortages resulting from load forecast error on the shoulders of capacity suppliers, as capacity resources not committed in the day-ahead market that could not come on-line in time to meet load or provide serves during critical hours would forgo the capacity payment.

This responsibility has been modified in the FCM proposal with the elimination of performance penalties for resources that were offered but not committed in the day-ahead market,[136] but the financial incentive for LSEs to schedule expected load in the day-ahead market is limited by low levels of shortage pricing.

Maintenance Outages

A fourth issue is that while not proposed as part of the ISO-NE LICAP or FCM designs, a system of availability incentives tied to generating unit performance during real-time shortage conditions has the potential to simplify and improve other elements of capacity design, such as the managing of maintenance outages. Under UCAP systems generators are not penalized if they are unavailable during scheduled maintenance outages. To ensure that capacity is available when needed, maintenance outages must be approved by the ISO and the ISOs can require that maintenance schedules be adjusted if the ISO foresees a reliability problem. The burden of ensuring that maintenance outages are scheduled during low risk periods is in part shifted to the supplier under a capacity market system tied to availability during reserves shortages, as the penalty would be incurred regardless of why the unit is not available. Under such a system suppliers would still benefit from central ISO coordination of outages, because suppliers would not be

---

136 Settlement Agreement, Section, Item 11, V.C.4.a. The elimination of performance penalties is initially limited to resources with a notification and start-up time of 16 hours or less, with an eventual reduction to 12 hours.

aware of the outages plans of their competitors while the ISO would. A capacity market system tied to actual resource availability during reserve shortages, however, would provide improve incentives for suppliers to complete maintenance quickly if system conditions changed resulting in increased potential for reserve shortages. Under the current FCM proposal, capacity resources that are unavailable during reserve shortage hours as a result of a planned maintenance outage approved by the ISO would not suffer a performance penalty as long as the outage was scheduled on one of the shoulder months.[137]

Another area of tension related to maintenance outages under a UCAP type capacity market system that could potentially be ameliorated under a reserve shortage-based performance incentive system is the availability of quick start units that were not scheduled to either operate or provide reserves in the day-ahead market. Under many UCAP systems quick start units that are not scheduled in the day-ahead market are nevertheless required to be available in real-time, in case they are needed. Such a requirement means that the supplier incurs the cost of manning the units on all such days, which would raise the cost of providing capacity.[138] Another impact of these availability requirements is on preventative maintenance. The operational status of generating units does not always fall into polar categories of unable to operate and in perfect condition. Quickstart units may be available to operate if needed, but simultaneously have performance issues that it would be desirable to address with preventive maintenance if the units are not needed. Suppliers therefore like to be able to undertake preventive maintenance of quickstart units if the units are not scheduled to generate or provide reserves in the day-ahead market. Under a traditional UCAP system, supplier preferences regarding manning and maintenance decisions need to be constrained by the ISO as the suppliers do not bear the full consequences of being unavailable should system conditions change (i.e., they receive capacity payment and do not forgo substantial shortage-driven revenues in the real-time market).

Under an ISO-NE type capacity market system with payments tied to resource availability during reserve shortages, the system operator should be able to provide suppliers with more discretion regarding both manning and preventive maintenance decisions as the supplier will bear substantial costs if its unit is unavailable during a real-time shortage hour as a result of the supplier's decisions. The ISO may still need to provide a coordination function, however, as individual quickstart suppliers making manning and preventative maintenance decisions will not be aware of the similar

---

137 Settlement Agreement, Item 11, Section V.C.4.d.

138 Under a demand curve system when defining the cost of the marginal unit, the cost of these requirements needs to be included in the benchmark. It needs to be kept in mind that the cost of requirements imposed on ICAP [capacity] providers is ultimately borne by consumers, so imposing onerous requirements on capacity market suppliers is not free. In the event of a generator outage or transmission contingency would not be scheduled to provide reserves.

decisions being made by their competitors, but the potential for reserve shortages depends on these collective decisions.

Deliverability

A fifth issue is that the implementation of a reserve shortage hour availability metric could simplify deliverability assessments for capacity resources as resources that were not delivering during shortage conditions would not be dispatched for energy nor selected to provide reserves.[139] Units that were not deliverable would therefore forgo capacity payments for their lack of deliverability just as if they were unavailable due to a lack of gas or a forced outage. The interaction of deliverability and capacity availability metrics has some complications, however. A lack of deliverability by generation within a generation pocket does not mean that none of the generation in the load pocket is deliverable, it just means that some of the generation is not deliverable. Moreover, which generation is dispatched in real-time, and thus satisfies the capacity availability standard, is not necessarily the generation resource that previously satisfied some theoretical deliverability test, it is the generation resource within the load pocket with the lowest energy offer price (abstracting from differences in constraint impacts). If thousands of dollars per MW hour in capacity market payments rest on which units within a generation pocket are dispatched for energy the potentially affected generators therefore have an incentive to submit large negative offer prices in the energy dispatch or self-schedule to ensure that they are operating.

Consider first an installed capacity system including both an ex ante deliverability test and an ISO-NE type availability incentive based on performance during hours of reserve shortage. Under such a system, installed capacity resources would have large installed capacity payments riding on whether they were dispatched to provide energy or scheduled to provide reserves so capacity market resources located within potential generation pockets would have an incentive to submit energy offer prices low enough to ensure that they were dispatched in preference to non-capacity market resources within the pocket. If the capacity market resources were not the lowest cost resources from the standpoint of energy cost this would not be fully efficient but the actual impact would likely be very small given the likely small number of reserve shortage hours and the unlikelihood that substantial undeliverable capacity would be constructed.

Under such an availability metric even generators that have been found to be deliverable in a probabilistic ex ante deliverability test might not be dispatchable during a particular reserve shortage hour depending on the pattern of generation and transmission during the hour outages have reduced transfer capability and the resources are not

---

139 In systems with explicit reserve markets, security analysis is generally applied to reserve scheduling and resources (or capacity segments) that could not be dispatched in the event of a generator outage or transmission contingency would not be scheduled to provide reserves.

scheduled to provide either energy or reserves. In such a circumstance, the affected capacity market resources would have an incentive to submit substantially negative offer prices to ensure that they are dispatched in real-time, magnifying real-time congestion.[140] Under the FCM proposal, capacity resources that cannot be dispatched in real-time as a result of transmission congestion, whether due to transmission outages or normal congestion patterns, do not incur performance penalties as a result of their inability to be dispatched.[141]

Now consider an alternative system in which there is no ex ante determination of which resources are deliverable for capacity market purposes. Under this system capacity payments would flow to those resources that are actually dispatched in real-time. Generators could assure themselves of financial deliverability by purchasing FTRs sourced in the generation pocket and sinking in at load in the capacity zone. A generator hedged by holding such an FTR could capture the capacity market value in the day-ahead market by offering its generation at negative prices. Even if the generator were not scheduled in the day-ahead market, it would capture the value of the capacity market payment in the value of its FTR as long as the shortage conditions were anticipated in the day-ahead market. If a resource scheduled in the day-ahead market were displaced in real-time by a resource submitting even lower offer prices that drove the real-time price even more negative than day-ahead prices, the resource owner would be hedged by its day-ahead schedule. If the price at the source of the FTR owning resource's schedule were substantially negative in real-time and the resource did not operate, settling the difference between its day-ahead schedule at real-time output at real-time prices would entail a payment by the ISO to the resource for the output it did not deliver in real-time. FTRs and day-ahead schedules would therefore operate in a manner fairly similar to in an energy only market, except that the value of the FTRs in a generation pocket would be set by very negative prices inside the generation pocket and $1,000/MWh prices outside the load pocket. The capacity market returns of resources located in such generation pockets would be reduced by the cost of purchasing FTRs in the relevant FTR auction. If there were more capacity resources located within the generation pocket than the FTRs that could be awarded from sources within the pocket, the resources within the pocket would in effect earn a reduced capacity payment, the nominal capacity payment less the excess congestion charges in the day-ahead market.

---

140 These congestion rents would not flow to FTR holders because LICAP availability incentives are based on the real-time dispatch not schedules in the day-ahead market. If the reserve shortage conditions were anticipated, capacity market resources would seek to schedule their generation in the day-ahead market in which case the capacity market congestion rents would flow to the FTR holder.

141 Settlement Agreement, Item 11, Section V.C.4.b.

Energy Limited Units

A sixth issue is that a capacity market with performance incentives based on availability during reserve shortages is fundamentally different from an energy pricing system in terms of defining the supply obligations of energy limited units. An important feature of capacity market performance incentives is that the annual or monthly availability requirement of capacity resources is defined by the number of hours of reserve shortage, not by terms specific to the resource. Whether a limited energy resource will be able to satisfy the availability requirement and avoid excessive penalties during a period would depend on the number of reserve shortage hours during the period compared to the energy limits of the resource.[142] The number of reserve shortage hours, however, does not depend on the performance of the energy limited resource but on the aggregate performance of the other resources in the capacity market, the level of imports during the hours of potential reserve shortage, the accuracy of the ISO's peak load forecast, and the accuracy of the ISO's estimate of the load shape.[143] Under an energy-only pricing system, an energy-limited resource could contract with an LSE to supply energy during a defined number of high cost hours and the LSE would take responsibility for ensuring that it contracted for enough baseload, intermediate and peaking capacity to meet the shape of its load over the month and year.

The concern with the potential effects of performance incentives or energy limited resources is not that energy limited resources would potentially earn lower returns in a capacity market with performance incentives than would similar resources that are not energy limited. Such an outcome would be efficient and the same outcome would prevail in an energy only market, an LSE would pay more for a call contract good during 8,000 hours than a call contract good during 200 hours. Rather, the issue is the shift in responsibility for analyzing the effectiveness of the resource mix for meeting load. Under an energy-only pricing system the LSE would analyze and contract for the mix of resources required to meet its load shape, just as the vertically integrated utility did historically. Under a capacity market system with performance standards, however, this burden implicitly shifts to the resource owner which in formulating its offer price would need to analyze the overall adequacy of the resource mix in order to evaluate the number of hours it would potentially be called upon to operate and its ability to actually earn its capacity payment. Aside from the unnecessary complexity imposed on limited energy

---

142 It would continue to be important in such a system to not subject the offer prices of the energy-limited units to offer price mitigation, as this could cause the resources to be dispatched in non-shortage hours, impacting reliability and potentially subjecting the resources to capacity market performance penalties if they were unavailable during actual reserve shortage conditions.

143 In addition, under either a capacity market or energy only resource adequacy mechanism, it is important that the energy offer prices of energy limited units not be subject to cost based offer price mitigation as offer price mitigation could prevent the resources from submitting high enough offer prices to avoid being dispatched in other than reserve shortage conditions, exhausting their annual operating hours during not critical conditions.

resources in formulating offer prices, performance incentives structured in this manner could potentially reduce the supply of capacity in a capacity short environment. If load has been under forecasted or if insufficient capacity has been offered in the forward capacity market, limited energy resources would have to anticipate more than the normal number of reserve shortage hours and a potential for more hours than the resource was capable of supplying, perhaps as a result of environmental permit conditions. The limited energy resource might therefore expect to retain too little of capacity payment for operation to be profitable, precisely because the market is in a shortage situation.

Energy and Reserves

A seventh issue also concerns energy limited resources but relates to short-term energy limits rather than monthly or annual energy limits. During an operating hour and day during which there is no shortage of capacity capable of providing operating reserves and thus able to operate for short periods of time in the event of generation or transmission contingency but there is a shortage of resources able to generate energy on a sustained basis over the day, the price of energy might be very high, but the price of reserves would be relatively low. This circumstance is readily addressed in an energy only market design as there are separate prices for energy and reserves. Both the energy and reserve price would be high during periods of capacity, i.e., reserve shortages, but only the energy price would be high during periods of energy shortage. These prices would provide efficient incentives for resources to be able to supply both energy and capacity in the long-run and for the scheduling of resources between reserves and generation during the operating day.

A capacity market design in which resources earn the capacity price for being available during hours of reserve shortage, may not handle energy shortages as well, however. Resources will bid to supply contingency reserves at low prices, but would not be available to generate energy on a sustained basis. If the day-ahead market worked effectively, sufficient resources would be scheduled day-ahead to meet the expected load, at appropriately high day-ahead prices, given the availability of fuel. If real-time load turns out to be lower than expected day-ahead, gas and power prices will likely decline and there will be no reliability issue. Suppose, however, that actual weather is colder than anticipated and intra-day gas prices are higher than day-ahead gas prices and additional energy beyond that scheduled in the day-ahead market is needed to meet load. If offer price mitigation is applied based on the prior days gas prices which are lower than actual intra-day prices, generators will be unwilling to buy gas at high intra-day prices if they will not be able to recover the purchase cost. In addition, since resources will likely need to purchase the additional gas in the intra-day market before knowing exactly how they will be dispatched, suppliers would not find it economic to buy this gas unless they expected to be able to earn more than their incremental costs if they were in fact dispatched.

Overall, performance incentives for capacity market resources tied to availability of the resources during reserves shortage conditions have limitations relative to the incentives provided by an energy only marker but nevertheless have the potential to provide better incentives than the UCAP methodologies currently in use.

F. Capacity Imports

A further element of a capacity market system is the need to account for imports and exports. This has two aspects. First, a critical component of all Eastern capacity market systems is the right of recall during shortage conditions for exports supported by capacity market resources. In the NYISO, external transactions supported by capacity market resources are subject to real-time curtailment to resolve a NYISO reserves shortage.[144] In PJM, all exports supported by capacity resources may be interrupted to serve PJM load if PJM declares a maximum generation emergency.[145] Similarly, sales out of the New England control area from capacity market resources can be interrupted to serve New England load if ISO-New England declares an emergency condition.[146] These recall provisions can create seams and price differentials during shortage conditions across multiple control areas but do not interfere with the scheduling of imports and exports to minimize the cost of meeting regional load under non-shortage conditions.[147]

The second element of these rules is the treatment of capacity imports. Traditionally, the Eastern power pools assumed that a certain amount of power would be available on some interface under stressed conditions and netted this from the collective pool capacity requirement. To avoid double counting of the import power relied upon in this reliability analysis, PJM for example, imposed a CBM margin which made most of the external transfer capability unavailable to support firm imports.[148] With the development of explicit recall rules, this logic is less compelling as external capacity market resources must be dedicated non-recallable capacity.

---

144 NYISO Services Tariff, Section 5.12.10.

145 PJM Operating Agreement, Schedule 1, Section 1.11.3A, June 22, 2005.

146 ISO New England, Market Rule 1, Section III.1.11.4, December 22, 2004.

147 Stoft and Cramton 2006 (p. 56) propose that California forbid exports supported by capacity resources during high load seasons without regard to whether shortage conditions exist. Such a provision would raise the cost of meeting regional load, and reduce the energy market revenues of generation located in California, raising the required capacity price. Since the effect of such a rule would be to require that capacity located in California remain undispatched at the same time that energy prices outside California exceed the cost of that capacity, it is difficult to envision FERC approving such a "Must Not Offer" rule.

148 The PJM CBM margin was quite different from CBM margins used in the Midwest. The PJM CBM margin only reserved the sale of this capacity as firm transmission service, thus making it unavailable to support firm capacity imports. In real-time all of the transfer capability was made available for use to support non-firm transmission. The PJM CBM margin was simply a mechanism for managing reserve requirements. There was no need to restrict use of CBM capacity to support imports in real-time as these imports met PJM load just as well as emergency imports. See PJM OATT, Attachment C, Sheet 280.

The NYISO places an overall limitation on the amount of the capacity requirement that can be met with external resources (2,755 MW) and then places additional interface by interface limits on external resource capacity market resources.[149]

ISO New England also makes the transfer capability of each interface, net of grand fathered agreements and less any tie line benefits assumed in calculating the capacity requirements, available to support capacity market imports.[150]

One problem that has manifested itself with respect to capacity imports is the need for a different approach to security analysis of import capability, as in some cases transmission maintenance outages can dramatically reduce inter-control area transfer capability and render the output of external capacity resources undeliverable. Thus, it is probably not appropriate to make the entire N-1 transfer capability on an external interface available to support capacity imports, as a single maintenance outage could make much of this capacity unavailable. There may therefore be a need to shift toward using N-2 transfer capability to define transmission limits for imports of external capacity market resources.

A second contentious issue is the treatment of units seeking to split their capacity between pools. The ISO-NE UCAP rules permitted potential capacity market suppliers to delist their resources as qualified NEPOOL capacity resources prior to the start of the obligation month effective at the beginning of the month.[151] An important and controversial element of the ISO-NE capacity market rules was that only whole resources could be delisted.[152] One reason for such a rule is to facilitate monitoring of compliance with outage and derating rules, which are potentially subject to circumvention if the market participate can choose how to assign outages between multiple capacity markets. Split capacity market units also complicate market power mitigation and enforcement of the DAM bidding requirement, as software needs to account for distinct physical unit upper limits and upper limits committed to a particular capacity market.

A third issue relating to imports is methodology for assessing the amount of imported energy that will be available under stressed system conditions. This has been addressed in the Eastern capacity markets by including external adjacent regions in the Monte Carlo analyses used to develop the capacity requirement. The reliability analysis therefore models transmission constraints in the external regions and accounts for the potential for correlations in weather conditions that will limit the ability to rely on external supplies. For example, the New York State Reliability Council models the outside world by matching their three highest peak load days to the corresponding

---

149 NYISO Installed Capacity Manual, Section 2.7, and Attachment B.

150 NEPOOL Manual for ICAP, Attachment G, Section 1.5.

151 ISO-NE Market Rule 1, Section 8.3.4, Sheet 86; NEPOOL ICAP Manual, Section_3.9.1.

152 ISO-NE Market Rule 1, Section 8.3.4

NYISO load levels.[153] PJM's determination of its reserve margin includes modeling of the rest of the world, including NPCC, SERC and ECAR.[154]

In terms of the impact of a capacity market requirement on energy prices, its impact can be very different depending on whether the capacity market region is an importer during periods of shortage conditions in adjacent markets. If there are no reserve shortages in the adjacent control areas, then a capacity market region will be able to attract imports at prices less than or equal to its price cap, because its offer price would determine the price of power in the adjacent regions. The supply of energy would be limited by the supply surplus in the adjacent control areas, not the price cap. It is therefore important in modeling the supply of imports to assess whether the conditions producing the reserve shortage conditions within the capacity market control area would also impact adjacent control areas. For reserves shortage attributable to unusually large outages, such as the outages of multiple nuclear plants, these events would likely be uncorrelated across the control areas so large imports could be available. For reserve shortages attributable to weather conditions, however, there can be a degree of correlation between the conditions within adjacent control areas that limit the level of imports. Between New England and NYISO, for example, a very hot and humid summer day in New York City may also be a very hot and humid day in Connecticut and Boston and this needs to be factored into the control area analysis of the supply of imports.

This is a particularly important issue in the West because one source of reserve and energy shortage conditions is the hydro cycle, which may produce correlated reductions in supply across several control areas. During low hydro years there may be a number of control areas that are reserve short net importers and the price of power may be determined by the offer price of reserve short control areas whose offer price reflects the value of lost load or the cost of entering an emergency state, rather than the price cap in control areas relying on capacity markets. In this situation, the control area relying on a capacity market for reliability will likely find that imposing a price cap merely ensures that its consumers bear all of the reserve shortages in the WECC, which is what happened to California in 2000-2001. If the capacity market region needs to be a net importer of energy during shortage conditions in adjacent control areas, then the energy market price must rise to the level in the adjacent control areas in order to attract imports. If the adjacent control areas do not rely on a capacity market system for reliability and have a higher energy market price cap than the region using a capacity market, then imports may be available during widespread shortage conditions only at prices in excess of the price cap in the capacity market region.

A critical issue for a capacity market system in California and other western control areas is therefore what is assumed in terms of import supply during shortage

---

153 NYSRC 2004, p. 35.

154 PJM Manual 20, p. 19.

conditions. If western control areas plan to meet their firm load in part with imported power during low hydro years, the market design needs to anticipate very high energy prices during those years. Unless some sort of disconnect is enforced between the price paid for imported energy and energy supplied by internal generation, the market design should anticipate that internal energy prices will rise to the level of prices outside the capacity market control area.

If the capacity market region is a net exporter during the shortage conditions, then these exports can be recalled at prices no higher than the bid cap in capacity market region. Depending on details of the pricing system, exports may be recalled at prices well below the price cap. For example, on May 8, 2000 when PJM had recalled all scheduled imports to New York and was also buying emergency energy from New York, the PJM energy market price never rose above $483/MWh despite the $1,000/MWh bid cap.

G. Rate Stability, Forward Contracting and Demand Response

Several of the issues relating to choices between reliance on energy only and capacity markets to support resource adequacy have touched upon the role of forward hedging. This section returns to those topics and considers several questions relating to the role of forward hedging, rate stability and demand response.

First, an important motivation for relying on capacity rather than energy markets to support resource adequacy is a belief that a relatively level set of capacity payments derived from an explicit capacity market will be subject to less regulatory risk than implicit capacity payments derived from margins in the energy market during shortage conditions.[155] As discussed above, energy only markets are likely to produce capacity returns that are concentrated in particular years which may result in more volatility in consumer costs than would be permitted to occur in existing regulatory structures. This volatility could be avoided through forward hedging of energy costs, but retail access has limited the extent of forward hedging of either energy or capacity charges by LSEs. If the return of fixed operating costs and return on investment is concentrated in a few tight supply years in an energy only market system and capacity suppliers do not expect that prices would actually be allowed to rise to the required level, then an energy-only market will in practice not support the intended level of capacity. This could potentially become a self-fulfilling prophecy as if suppliers do not expect high prices to be allowed to prevail and therefore do not build the level of capacity required to avoid large numbers of shortage hours during tight supply years, the power market will not appear to be

---

155 This hypothetical comparison assumes a comparable level of payments over time. Inefficiently high power costs will also give rise to regulatory risk, whether the high costs are level of not, as has been pointed out by National Grid, among others. See Initial Brief of National Grid USA, Docket No. ER03-563-030, April 15, 2005, pp. 25-29.

operating efficiently, justifying the regulatory intervention that is actually the source of the problem.

This rationale for reliance on capacity markets to support resource adequacy raises the further question of whether the capacity payments derived from an explicit capacity market are actually more stable and less subject to regulatory risk than implicit capacity payments derived from energy markets. While there is regulatory risk associated with high power prices for unhedged retail customers during shortage years, there is also regulatory risk associated with capacity market charges for loads that remain high or even rise during periods of capacity surplus. As discussed in Section III.C, actual experience has been that when the capacity markets in PJM and ISO-NE rose toward the level required to support the entry of new capacity, the increase was asserted to arise from the exercise of market power and actions were taken to reduce capacity prices. Moreover, a central feature of the FCM proposal is that the actual application of the locational capacity market and capacity payments is postponed into the future following a transition period with specified levels of capacity payments and no locational component to the payments.[156]

In practice the NYISO capacity markets are the only capacity markets in the U.S. that have actually resulted in more than token capacity payments on a continuing basis.[157] It has yet to be seen whether capacity markets can be implemented in PJM or ISO-NE in a manner that actually results in material capacity payments to resources located in supply deficit areas and provides sufficient payments to maintain resource adequacy on a continuing basis.

Second, the introduction of a capacity market implicitly hedges consumers against capacity charges by reducing the potential for extreme capacity payments arising from sustained reserve shortage conditions. In addition, accompanying low levels of shortage pricing and/or bid and price caps suppress prices during reserve shortage conditions. If the capacity market design includes an after the fact adjustment for peak energy market rentals as in the ISO-NE LICAP and FCM designs, consumers are largely hedged against the shortage cost component of energy prices. On the other hand, capacity markets do not hedge consumers against increases in the cost of power due to gas price increases or NOx allowance cost increases. This is the case both for the existing capacity market designs in NYISO and PJM and the proposals for LICAP and FCM in ISO-NE.

We have seen during 2005 and 2006 that large changes in gas prices can produce political and regulatory problems within capacity market systems, with large increases in cost of power resulting in large increases in the default or provider of last resort rates paid

---

[156] Settlement Agreement, Item 11, Section VIII.

[157] Whether these payments are viewed as too high or not high enough, they are unambiguously more than token, even for capacity outside New York City.

by many or most retail customers. Most of the increases in the price of power in the west during 2000-2001 were associated with increases in the price of natural gas and NOx allowances rather than reserve shortage conditions, so would not have been directly hedged by a capacity market system. Not only would these price increases not have been hedged under a UCAP-type system, they also would not have been hedged under a capacity market system including a deduction for peak energy rentals due to reserve shortage conditions. Thus, it needs to be recognized that implementation of capacity markets does not by itself solve the political problem of substantial year to year changes in power costs. The potential for large year to year variations in power costs needs to be addressed with an efficient structure of forward energy contracts.

Third, forward energy contracts are a better way to hedge consumer costs than capacity markets, forward energy contracts can potentially hedge power prices against changes in gas and emission allowance costs as well as capacity shortages. Regulatory policies that mandate forward hedging of either capacity or energy costs are not necessarily in the interests of consumers, however. As pointed out in the discussion of whether California should mandate forward hedging of power costs back in the fall of 2000, forward hedging needs to be based on consumer demands, it will be inefficient and problematic from a political and regulatory standpoint to mandate forward hedges at prices that exceed what consumers are willing to pay for that power.[158]

While it is commonly asserted that consumer demand is almost completely price inelastic, this is untrue in the case of sustained price prices. A very important element of the response to the western power crisis were the substantial reductions in power consumption by power intensive industrial customers, aluminum producers being a conspicuous example.[159] Moreover, when the California Public Utilities Commission finally raised retail rates in the late spring of 2001, the drop in consumer demand was dramatic.[160] The western hydro cycle appears to provide a good example of the kind of shocks to the system that would be more efficiently addressed with less adverse impact on reliability under an energy-only than capacity based resource adequacy mechanism. It would likely be extremely expensive for western utilities to construct sufficient non-hydro generating capacity, with a supporting fuel supply system (such as gas pipelines for gas-fired generation), to meet load during low hydro years without any demand side conservation in response to high prices. The capacity payments to hydro and other

---

158 See Chandley, John, Scott Harvey and William Hogan, "Electricity Market Reform in California," November 22, 2000, p. 9.

159 BPA press releases detailing these actions can be found at www.bpa.gov.

160 See, for example, California Energy Commission, 2002-2012 Electricity outlook Report, February 2002 pp. 1-9 and 1-10; Charles Goldman, Joseph Eto, and Galen Barbose, "California Customer Load Reductions during the Electricity Crisis: Did they help to keep the Lights on," May 2002. It is possible that part of this drop in consumption was a result of public spirited response to the crisis, but the crisis had existed since the summer of 2000.

limited energy resources might go to zero as a result of performance penalties arising from a very large number of reserve/energy shortage hours but this would not lead to supply demand balance without increases in the price of energy to consumers. Under an energy only market design, even load that is fully hedged against price increases would have an incentive to reduce consumption by shutting down operations if the price became high enough. It is not clear that any entity has the required incentives to take these actions under capacity market designs with performance incentives nor whether the requisite property rights of load to power would even exist to support reductions in consumption.

Thus, while it is sometimes suggested that consumers need to be fully hedged against the higher prices arising from reserve shortage conditions,[161] this is not necessarily desirable. It is important to avoid the outcomes in the 1970s and 1980s when capacity costs were locked in for power demand that did not exist at the prices required to cover those capacity costs. This requires that any process of forward contracting for energy or capacity include a comparison of the overall cost of power to the value of that power to the incremental power consumer, rather than simply contracting to buy power to meet a forecast level of power consumption that is derived without regard to the price of power.[162]

Setting up an autopilot forward procurement process for capacity and or energy without regard to the price of the power and the value of this power to consumers risks repeating the process of creating stranded costs.[163] The restructuring of the power industry in the Northeast, parts of the Midwest and in California was motivated in significant part by an intention to avoid such disconnects in the future. The design and implementation of retail access programs to date has generally not effectively addressed this problem, but it may be better to address the problems with retail access designs, such as the changes recently implemented by the New York PSC (discussed in Section III.C above), than to contract for power on behalf of consumers without regard to the value of that power to these consumers. These concerns are relevant both from the standpoint of overall economic efficiency and in assessing regulatory risk under forward capacity market systems, particularly since it is not clear how responsibility for capacity will be allocated if the capacity needed in the operating year is less than forecast three years in advance. Are the costs of this capacity implicitly subject to being stranded? Will consumers have to bear the cost of capacity contracted for by the ISO to meet demand that does not exist?

---

161 Cramton and Stoft (2006), pp. 5, 12, 49.

162 These kinds of issues were raised in the ISO-NE LICAP hearing by National Grid, among others. See Initial Brief of National Grid USA, pp. 4-5.

163 In the West, the most obvious risk is of contracting for capacity to meet load during low hydro years whose cost exceeds the value of marginal power consumption. In the East, the most apparent risk is of contracting for capacity to meet future load that never materializes because of continued high gas and power prices.

Fourth, a fundamental limitation of capacity markets is that because real-time energy prices are suppressed and do not reflect the actual costs of reserve shortage conditions, real-time loads do not have efficient incentives to reduce consumption during such conditions. Most capacity market systems attempt, with varying degrees of success, to attempt to replicate for consumers the incentives that would be present if customers saw real-time prices at the margin but these efforts are compromised by the capacity market structure. Capacity market systems can endeavor to replicate the incentives of an energy only market design by paying consumers not to consume but they inevitably provide too little demand response at too a high a cost because of the impossibility of accurately measuring what the consumer would otherwise have consumed.

Fifth, while a limitation of forward capacity market is that it only hedges future capacity, not energy costs, there are risks associated with proposals that would require more complete forward hedges of power costs. Forward contracts with power sellers will not be able to buck long-term changes in the overall energy market. Forward hedging of power contracts with gas-fired resource owners can incent gas pipeline construction, duel fuel capability and other measures that will ensure that local gas prices do not get too far out of line with Henry Hub on a long-term basis. Short-term volatility in the price of gas can be hedged to a degree in the futures market.[164] Power consumers, however, have a limited ability to hedge themselves against a sustain national increase in the price of gas. If the price of gas at Henry hub goes up on a long-term basis, the price paid by consumers will have to rise.

Trying to hedge consumers against sustained rises in the price of gas will likely be expensive for consumers, and consumers might find that in the end they are actually not hedged as the suppliers default. An optimal hedging strategy may therefore entail hedging a portion of demand with long-term purchases from resources with costs that are not tied to the price of gas (such as nuclear or coal generation),[165] locking in another portion of consumption with gas fired generators on an annual basis, and perhaps locking in capacity charges but not fuel costs for an additional portion of demand through long-term tolling agreements. In all of this hedging, attention would need to be paid to the actual load shape of demand over the year, and to the long-term and short-term price elasticity of demand.

---

[164] It needs to be kept in mind, however, that if consumers representing a material proportion of gas consumption attempted to hedge themselves in the forward gas market more than a few months out, the consumers might find the price of hedges rising.

[165] An underlying question is whether there are any imperfections in power markets that would raise the energy price risk of building nuclear and low emission coal generation resources that would be addressed through forward contracting by consumers.

H. Market Power

1. Economic Withholding in Capacity Markets

Although resource adequacy is often tied to market power mitigation, the implementation of capacity market does not mitigate the exercise of locational market power. If a resource owner has locational market power in energy markets, then it will also have market power in capacity markets. The ability of LSEs to enter into long-term contracts with generation entrants at competitive prices will often preclude or constrain the exercise of market power in long-term capacity markets, but long-term power contracts with entrants provide the same competitive pressure in long-term energy markets. Conversely, any potential for the exercise of locational market power that exists in spot energy markets would generally also exist in short-term capacity markets.

The most important difference between capacity and energy markets from the standpoint of market power is that economic withholding becomes progressively more difficult to identify as the timeframe moves further away from real time. In a centrally dispatched system such as PJM or New York, generation that is available (taking account of ramp constraints, deratings and environmental limits) and not generating energy or providing reserves in real time despite market prices that exceed its incremental opportunity costs can be identified as economically withheld. While the application of this criterion can be complex for units managing energy or fuel limits and during periods of volatile gas prices, it is generally possible to unambiguously identify substantial economic withholding. When one moves to the day-ahead timeframe, economic withholding is somewhat less clear cut as a competitive seller would not offer to sell power in the day-ahead market for less than the expected real-time price, regardless of its incremental costs, and the expected real-time price is not observable. Market participants expecting high real-time prices, however, can use virtual load bids to arbitrage any difference between day-ahead and expected real-time prices while making all of their capacity available for commitment in the day-ahead market and available for dispatch in real time.

In capacity markets, it is harder to define economic and physical withholding. In a capacity market, the long-run floor on capacity prices is provided by the avoidable costs of a unit that would not be recovered in energy and reserve margins. The avoidable costs of a unit can be roughly estimated based on historic costs as can past energy and reserve margins, but these past margins are not necessarily a good measure of current expectations. The shorter the timeframe for which the capacity payments apply, the harder it can be to distinguish economic or physical withholding from an unwillingness to keep money-losing capacity available. A capacity owner might keep a losing unit available for a period of days despite zero daily capacity prices but it would not agree to a forward commitment to keep the unit available for a sustained period of time as a capacity resource for a zero capacity price.[166]

The PJM market monitor concluded that the high capacity prices in PJM in early 2001 arose from an exercise of market power and the rules regarding the allocation of deficiency charges were changed to modify withholding incentives.[167] The market monitor's analysis focused in part on comparing the value of the recall right to the price of capacity, concluding that the offer prices in the daily capacity market were well above the value of the recall right.[168] While the value of the recall right should place a floor under the level of capacity prices in a competitive market, this level is not a ceiling. Energy prices could move in sync throughout the region producing zero differentials in energy prices between markets, yet generating capacity could require a capacity payment to remain in operation and provide capacity. This is one of the disconnects between the daily pricing of capacity and the term character of capacity decisions that complicates market power analysis as well as the effective functioning of capacity markets.

In this instance, a single entity was apparently in the position of being a pivotal supplier in the daily capacity market, which is consistent with the exercise of market power. On the other hand, the cost of making capacity available for a single day might exceed the PJM deficiency charge if the offered capacity was mothballed, which is not clear from the market monitor's discussion. While the daily capacity price was high during January through March, it was zero most of the rest of the year.[169] The competitive level of a daily capacity price is very poorly defined, as observed above, making it difficult to assess whether market power is being exercised. It is not the case, however, that market power is being exercised any time the capacity price exceeds zero.

The issues involving market power in short-run capacity markets extend not only to generators but also to retail access providers. Suppose that all generators have sold their capacity forward and all LSEs have purchased enough capacity credits to satisfy their obligations at prices ranging from $40,000 to $60,000/MW year depending on the date and duration of the purchase. Then suppose that LSE A loses 25 MW of load to LSE B. LSE A now has 25MW of extra capacity credits and LSE B has 25 MW too few. At what price should LSE A offer to sell its extra credits to LSE B. Retail competition would be chilled if LSE's losing load could set a price on their excess capacity credits

---

166 As noted above, the capacity price is also bounded by the price at which the resources capability could be sold for in adjacent regions or the value of being able to sell non-recallable power into adjacent regions.

167 PJM initially allocated deficiency revenues to holders of unsold capacity resources. Thus, capacity that was withheld from the market received the deficiency payments. This rule was modified effective June 1, 2001. See PJM Market Monitoring Unit, PJM Interconnection State of the Market Report 2001, pp. 69-70, 79-94 (hereafter PJM 2001).

168 PJM 2001, pp. 84-85. See also Anna Creti and Natalia Fabra, "Capacity Markets for Electricity," CSEM Working Paper, February 2004, p. 7.

169 See Figure 12 above, and PJM 2001, Table 2, p. 80.

reflecting the deficiency penalty because no additional supply would be available from capacity resources in the near term. Conversely, forward contracting for capacity by LSEs would be chilled if short-term transfers of capacity credits were always at prices close to zero while forward contracts reflected the actual cost of capacity.

An incidental feature of the ISO-NE FCM design is that these inter-LSE capacity credit transfers (and the need to set a price) are avoided because capacity is bought forward by the ISO and all LSEs are charged the forward price, not a short-term bilateral price.

New England has been the locus of concerns regarding the potential for the exercise of "market power" in exports, i.e., the scheduling of uneconomic exports. These concerns have manifested themselves in capacity markets with provisions intended to make it unprofitable for capacity suppliers to uneconomically export capacity in order to raise capacity market prices in New England.[170] There is, however, no potential for the exercise of market power through uneconomic exports. While a market participant could raise capacity prices in New England through uneconomic capacity exports to New York if there were a regulation that prevented capacity exports from New York to New England, there is no such rule. If there were a recognized potential for capacity prices in New England to rise above the value of capacity in New York, New York suppliers would offer capacity into the New England capacity market at the expected New York price. It is very basic in antitrust economics that the exercise of market power entails an ability to raise price on a sustained basis. This is inconsistent with theories of market power in capacity markets based on uneconomic exports.

Moreover, while there is a potential for LSEs purchasing capacity on a spot basis to be impacted by short-run surprise imbalances in exports and imports in the capacity market this potential exists only for those LSE that choose to purchase capacity in the spot market. Market participants that contract forward for capacity would be able to purchase capacity at the equilibrium price, without regard to short-run volatility in the capacity spot market, whether involving uneconomic exports or simply imperfect price convergence due to the nature of capacity markets. Rules that attempt to insulate LSEs from the consequences of buying capacity on a spot market basis do not address market power concerns, but simply distort contracting incentives and should be avoided.[171]

---

170 David LaPlante, Supplemental Direct Testimony, November 4, 2004 (hereafter LaPlante Supplemental), pp. 2-11; LaPlante Rebuttal, pp. 99-102; ISO-NE Initial Brief, Docket ER03-5763-030, pp. 72-73.

171 There is a more subtle problem involving exports that is unique to capacity markets, the treatment of exports sourced from a resource within a load pocket. In real-time or day-ahead LMP-based financial markets, the dispatch of a resource located within a load pocket to support an export, relieves congestion exactly as if the resource were dispatched to meet load within the load pocket. If the resource is dispatched, it relieves the constraint. It is not clear that the same assumption can be made in modeling exports by such a resource in the capacity market because there would be no obligation for the resource to offer its capacity in the real-time or day-ahead energy market.

Two important elements of the ISO-NE LICAP proposals were motivated by market power concerns; these were the rules treating existing capacity as fixed for the purpose of determining capacity market prices and the peak energy market rent adjustment for energy market revenues. The rules treating capacity as fixed for the purpose of determining the capacity market price were apparently motivated to address the potential for the exercise of market power through economic withholding by existing suppliers (including mothballed capacity) in the capacity market,[172] while the peak energy market rent adjustment was intended to eliminate any incentive to exercise market power through economic or physical withholding in energy markets.[173] Rather than addressing market power problems, however, these features were two of the more problematic features of the proposed ISO-NE LICAP design.

2. Offer Price Restrictions for Existing Capacity

As observed above, if suppliers possess locational market power in short-term energy markets, they will also possess market power in short-term capacity markets. To address this potential for the exercise of market power in locational capacity markets by incumbent resource owners, ISO-NE proposed to treat existing capacity (apparently including the capacity of new resources) as offered at a zero price for the purpose of clearing the capacity market.[174] FERC Staff[175] described this approach as altering the structure or design of the market in order to decrease market power in contrast to capping or resetting bids. Obviously, however, capping or resetting bids is exactly what the ISO-NE proposal would have done; it would effectively cap the offer prices of all existing capacity at zero for the purpose of clearing the market. Since the going forward costs of existing capacity are unambiguously non-zero, this market power mitigation approach has the potential to "clear" capacity that would not actually be available to met load adversely impacting reliability.

---

172 Initial Brief of ISO-New England Inc., Docket No. ER003-563-030, April 15, 2005, pp. 68-72. Steven Stoft, "Supplemental Direct Testimony," Docket ER03-563-030, November 4, 2004 (hereafter Stoft Supplemental 2004), pp. 3-8.

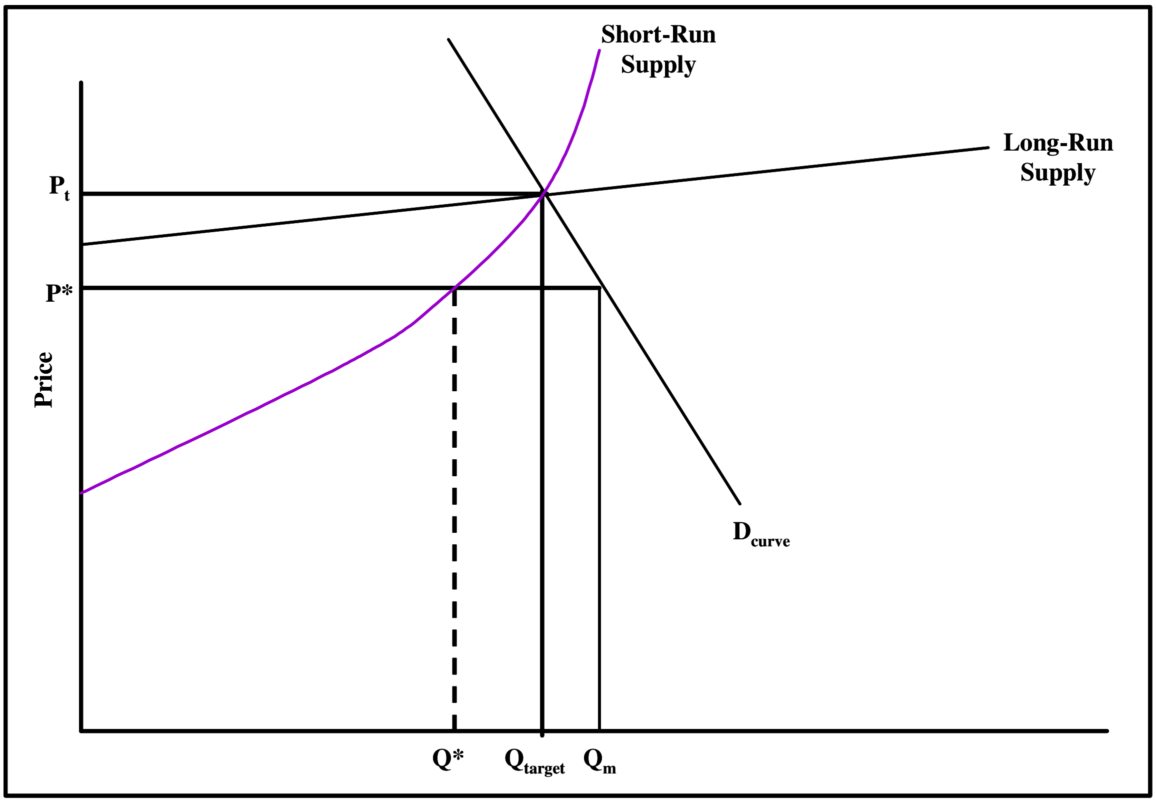
173 Stoft 2004, pp. 14, 19, 93-95; Cramton and Stoft 2006, p. 18.

174 See Stoft Supplemental 2004 pp. 4-8. LaPlante Supplemental, pp. 3-4, 12-13.

175 Initial Brief of Commission Trial Staff, pp. 33-34, April 15, 2005.

Suppose, for example, that the ISO estimated the long-run supply curve for capacity perfectly for the purpose of fixing the demand curve and that the short-run supply curve of capacity passes through the intersection of the demand curve and long-run supply curve as shown in Figure 32. The equilibrium price of capacity P_t would provide the target level of capacity Q_t, while capacity offered at prices above P_t (Q_m-Q_t) would be mothballed because the capacity payment would be insufficient to cover the going forward costs of keeping the capacity available.[176]

Figure 32
ISO-NE Capacity Prices with Shutdown Capacity



The ISO-NE market power mitigation rule would include this mothballed capacity in the supply (Q_m-Q*) used to clear the capacity market, so Q_m of capacity would be used to determine the capacity price (P*).[177] At the capacity price, P*, only Q* capacity would be available for operation, less than the target quantity for reliability purposes.

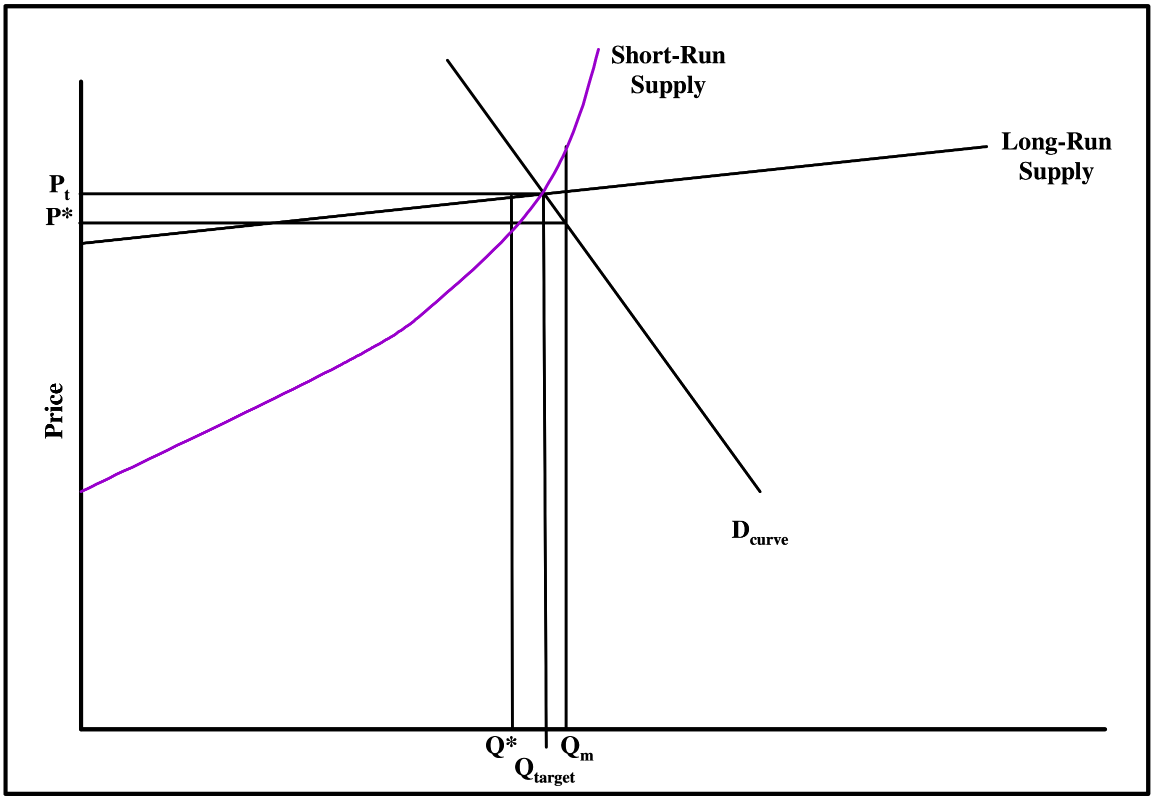
---

176 I.e., manned (not just manned when in operation, but having employees available to operate the plant when needed) in compliance with environmental regulations, and maintained in operating condition.

177 LaPlante Rebuttal, pp. 16, 99-100, 103-104, 107-108.

The capacity prices produced by this pricing rule would not sustain the target level of capacity and the supply of capacity would fall until the price determined by the pricing rule was consistent with the long-run supply curve as illustrated in Figure 33.

Figure 33
Equilibrium with Shutdown Capacity



This adverse reliability outcome would not arise under the ISO-NE LICAP market design if incumbent suppliers permanently shut-down high cost existing capacity as it became marginal, rather than offering the capacity in the capacity auction as the proposed rules included all active and mothballed resources in the capacity market demand curve but did not include resources that have been permanently shut down. This kind of behavior by incumbent capacity suppliers would not be desirable, however, as it could lead to spikes in the price of capacity when the incumbent suppliers misestimate the supply of new capacity and fail to offer capacity that would have been economic at the actual clearing price.

It can be foreseen that such rules would put existing suppliers in an awkward situation. If a supplier does not offer existing high cost capacity in the auction because the supplier does not expect the capacity to be economic in the capacity market, this decision will potentially cause price spikes in the capacity market if less new capacity is offered than expected or if the new capacity is offered at higher than expected prices. This kind of behavior might be hard to distinguish from physical withholding in the capacity market. Conversely, if the supplier offers its high-cost existing capacity in the auction and shuts the capacity down if the auction price is lower than needed to cover the going forward costs of the capacity resource, then the supplier will undermine reliability by shutting down the high cost capacity after the auction, leaving the ISO with less capacity than intended to meet reliability.

The apparent premise of this ISO-NE market power mitigation design was that mothballed capacity would be available when needed at zero cost in a capacity market design with low energy prices during shortage conditions.[178] Such an incentive for mothballed capacity to return to service during potential shortage conditions could exist in an energy only resource adequacy system with effective shortage pricing, but ISO-NE proposed to implement the LICAP capacity market precisely because the ISO-NE energy markets lack effective shortage pricing. If resources could recover their going forward costs in energy market revenues alone, there would be no need for a capacity market.

ISO-NE suggested that resources with offer prices in excess of the capacity price should be counted as capacity because they "make some contribution to reliability through their required participation in the real-time energy market and are physically located within the region's inventory of installed capacity."[179] This rationale assumes that the going forward costs of the resources are zero, so that the resources would be available in real-time even if they did not receive capacity payment. But if going forward costs are not zero, if they are actually substantial, then resources whose offer price exceeds the capacity price will be mothballed, will not be required to participate in either the day-ahead or real-time energy market, will not be available in real-time, and the lights will go out in New England. One might conjecture that if the capacity shortages were so extensive as to result in chronic rolling blackouts and high energy market margins with even modest shortage pricing, mothballed units might take steps to return to service. This outcome would not maintain the intended level of reliability because the costs would not return to service until the deteriorating reliability level produced increased energy market revenues. Worse, it would not happen at all under the ISO-NE capacity market design as the deduction for peak energy rentals discussed below would ensure that the combined capacity and energy market revenues remained inadequate to support bringing mothballed units back into service even with rolling blackouts.

Rather than ISO-NE allowing high cost existing resources that were treated as "cleared" in the proposed auction to shut down because the capacity price in the auction was less than the resource's actual offer price, one can anticipate that the ISO would seek to keep the resource in operation by entering into RMR contracts that would pay more

---

178 Initial Brief of ISO-NE, p. 72; Reply Brief of ISO-NE, pp. 117-118. FERC Staff also took this view, Initial Brief of Commission Trial Staff, p. 39.

179 Exhibit LIP-2, ISO-New England Inc.'s Responses to Trial Staff requests staff/ISO-NE 4-12 and Staff/ISO-NE 4-16. Similarly, see LaPlante Rebuttal, p. 108; Stoft Supplemental 2004, pp. 7-8.

than the clearing price in the auction in order to keep the existing capacity actually available. This RMR solution will be irresistibly attractive to entities with no long-term load serving obligations but would raise costs to consumers in the long-run. While the RMR contracts would appear beneficial to loads in the short-run by reducing the price paid for the low cost inframarginal existing capacity, consumers would be paying for existing capacity that is actually more expensive than some of the new capacity that was not purchased in the auction.

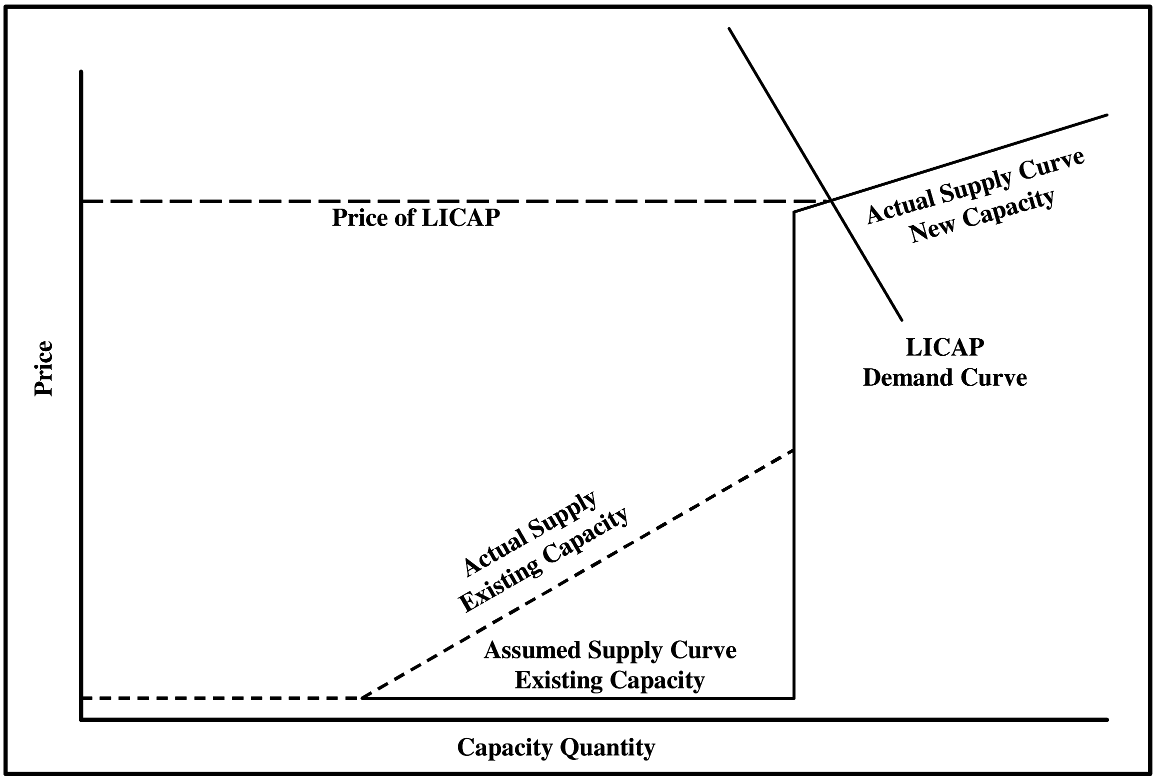
One of the problems with existing capacity markets that has been discussed above is that the capacity markets are so near term that most capacity costs are sunk at the time the market is cleared, so the capacity offer prices of new entrants could potentially reflect only a small portion of the actual cost of capacity on a long-term basis or could greatly exceed the cost on a long-term basis.[180] Forward markets for capacity, such as the CRAM, FCM and RPM proposals would likely permit offer prices in the capacity market to better reflect the full cost of new peaking units, as the construction costs of these units would not be sunk at the time of the auction and construction could be contingent on the resource clearing in the capacity market. The capacity offer prices of new capacity resources would be more closely related to the cost of new capacity if those costs were variable, rather than sunk, at the time that the capacity market is cleared and the capacity payment were fixed for multiple years.

---

180 If the new capacity resource is in the final stages of the construction at the time the capacity market is cleared, offering the capacity at its full cost could result in the capacity not clearing in the market, thus earning no capacity market revenues, yet few if any costs could be avoided if the capacity of the resource were not sold in the capacity market. Even the costs required to finish the plant might be effectively sunk by that time under the terms of the construction contract.

If the going forward costs of existing capacity are strictly less then the offer prices of new resources in the capacity market then the mitigation of offer prices from existing resources would not impact the capacity price or reserve adequacy. Thus, if forecast demand grew steadily, the clearing price for capacity would always be determined by the offer price of new capacity resources and would not depend on the actual offer price of existing capacity resources. This situation is portrayed in Figure 34.

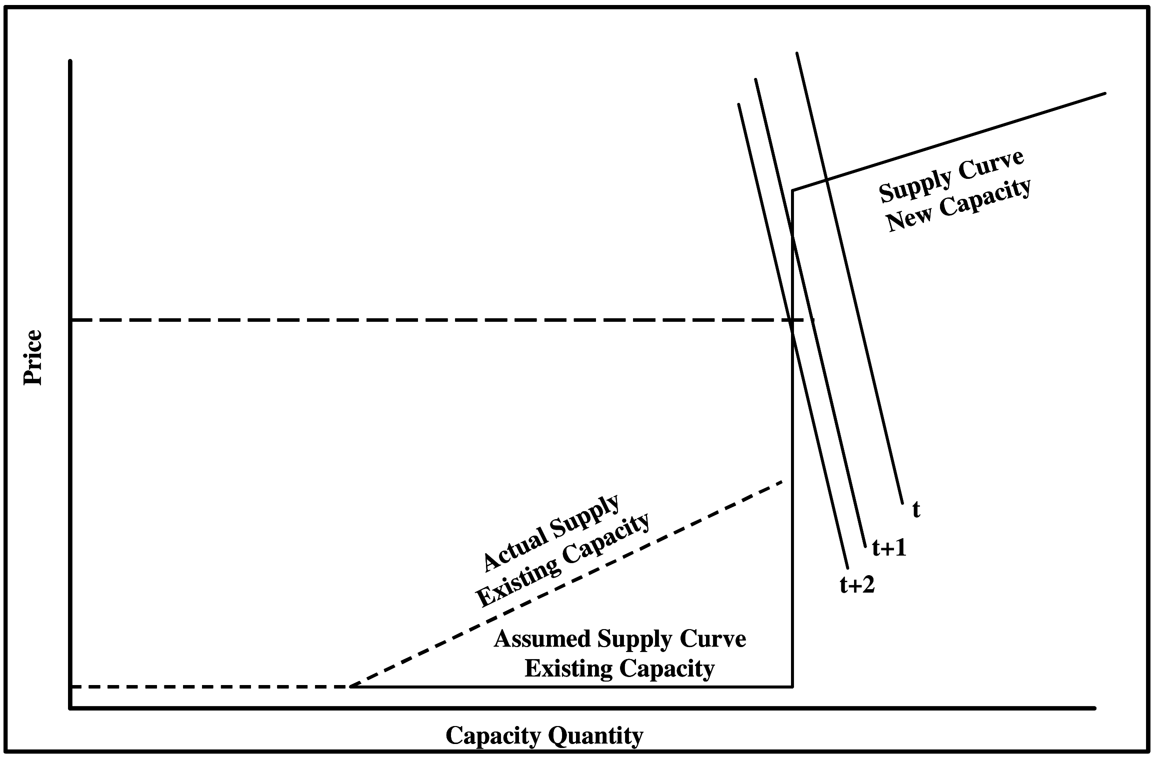
Figure 34
Capacity Market Prices with Growing Demand



If the capacity demand curve were shifting out from year to year and there were no excess capacity and the going-forward cost of existing capacity was always less than the full cost of new capacity, then the competitive price of capacity would be independent of the actual supply curve for existing capacity. In this circumstance, constraining existing capacity to be offered in the capacity market at a zero price would not distort capacity prices, but would prevent the owners of existing capacity resources from exercising market power by economically withholding capacity from the auction, because the capacity would not be marginal, even based on the actual supply curve.

Alternatively, suppose that the demand for capacity does not grow steadily from year to year but sometimes declines between successive years as portrayed in Figure 35,[181] perhaps because of changes in economic conditions. Figure 35 shows that over a considerable range of demand reductions, the capacity price still would not depend on the actual supply curve of existing capacity as the price would be set by the intersection of the sloping capacity demand curve with the discontinuity in the capacity supply curve.

Figure 35
Capacity Prices with Falling Peak Demand Forecasts



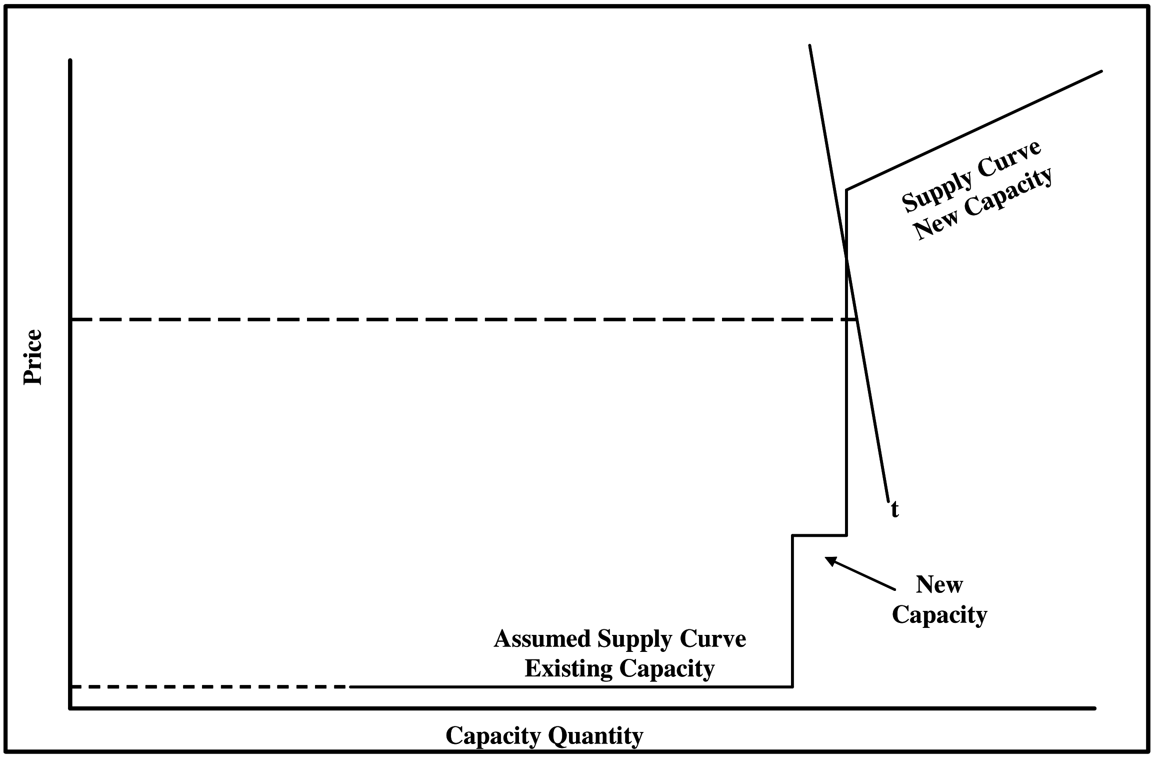
The capacity clearing price could also be determined by the demand curve when demand is growing over time as a result of new capacity that is offered at infra-marginal prices because its entry does not depend on capacity prices.

Not all new capacity, however, will be peaking units. Some new capacity might, for example, might be provided by new coal fired units whose construction costs would be largely sunk even at the time of a forward auction. These new resources might therefore offer their capacity into even a three year forward capacity market at relatively low prices as illustrated in Figure 36. These new long development time resources might completely eliminate the need to add any peaking units to meet the year to year growth in

181 The situation portrayed in Figure 35 could also arise in the context of a local capacity requirement from the construction of additional transfer capability into the local region.

demand, so that capacity price would be set by the demand curve, rather than the offer prices of new peaking units, even in a forward market with growing capacity requirements.

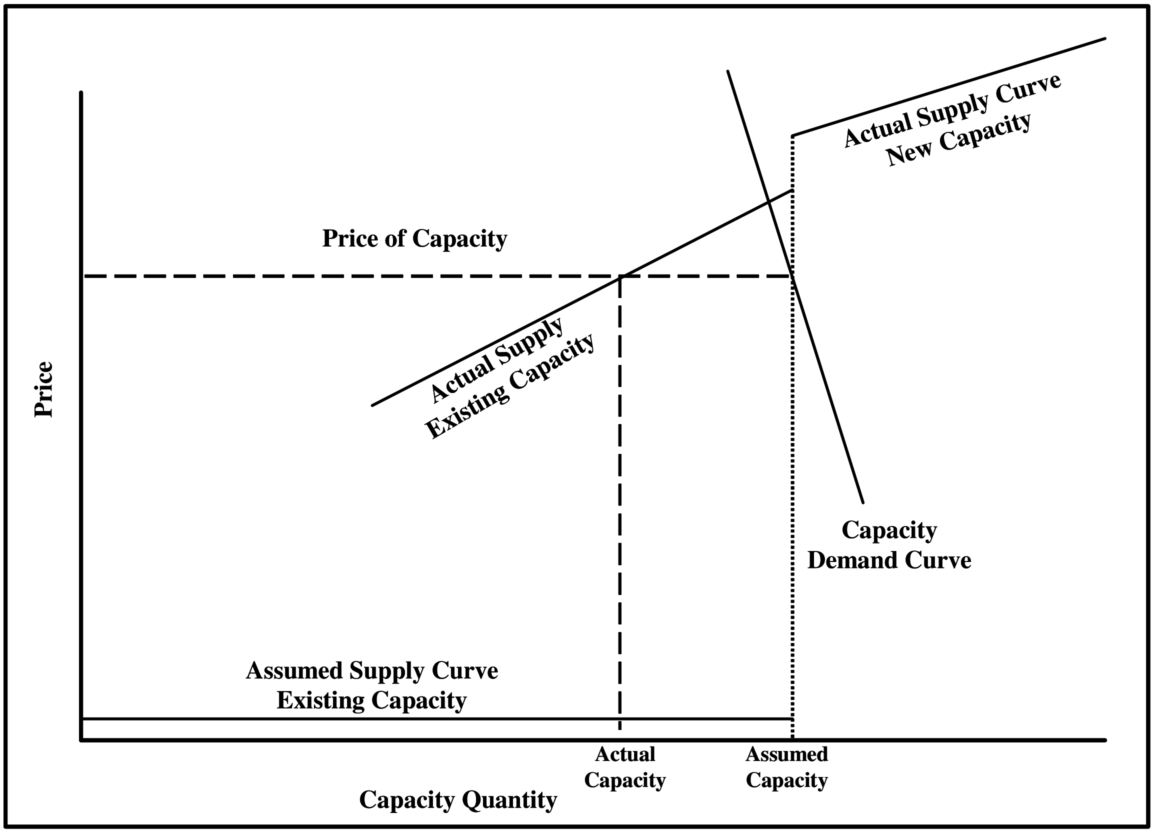
Figure 36
Capacity Prices with Inframarginal Entry



In this situation, the clearing price in the capacity market would not be determined by the offer prices of new peaking units, but would be determined by the intersection of the demand curve and the vertical portion of the supply curve, so the clearing price would be independent of the assumed offer price of the existing capacity.

An implicit assumption underlying the outcomes portrayed above is that the actual supply curve for existing capacity intersects the target capacity level well below the clearing price of capacity, whether the clearing price is determined by the offer prices of new capacity or by the demand curve. Suppose, however, that the actual supply curve for capacity is high enough relative to the clearing price of capacity that not all existing capacity is offered at prices below the clearing price in the auction, as illustrated in Figure 37. In this situation the application of the capacity demand curve would produce the same capacity market price as in Figure 35, but the assumed supply curve for existing capacity would in this situation overstate the capacity actually supplied as some of the existing capacity would shut down at the "clearing" price in the capacity market, and the forward capacity market produces a capacity shortfall relative to the reliability target.

Figure 37
Capacity Prices with Costly Capacity Market Obligations



This miscalculation would potentially have adverse reliability consequences as the amount of capacity available to meet load might be materially less than the target quantity needed to maintain reliability.

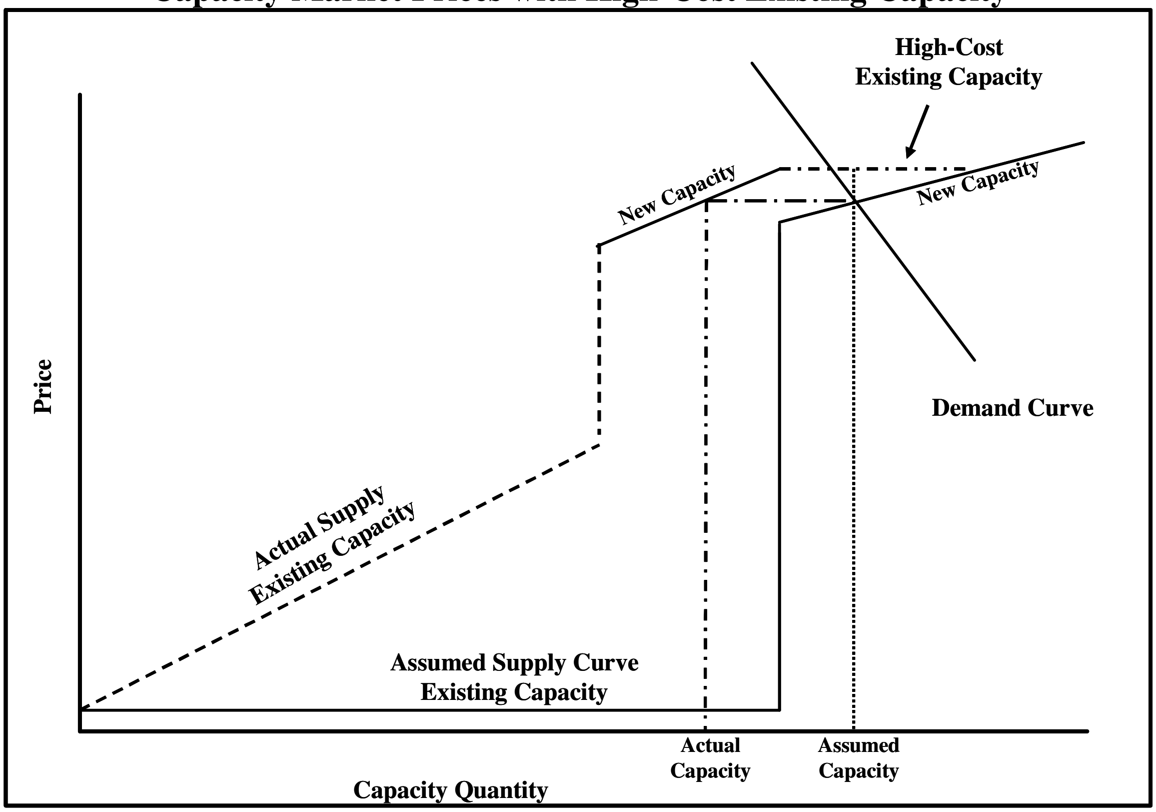
It would be desirable under a system in which existing capacity is treated as fixed in clearing the capacity market to avoid burdening capacity suppliers, and particularly existing capacity with costs that are not reflected in the determination of the target price for the demand curve as this would make this kind of outcome more likely.

The potential for such an approach of treating existing capacity as offered at a zero price to result in inadequate capacity even in a forward capacity market is unavoidable, however, and is not limited to the circumstance in which the price of capacity is determined by the demand curve, rather than the offer prices of existing capacity. While the implicit assumption that the going forward costs of existing capacity will always be less than the full cost of new capacity will generally be valid for the bulk of existing generating capacity, it is not plausible that it would be true for all existing capacity. As capacity ages there inevitably comes a point in time at which the construction and operation of new capacity is lower cost than maintaining existing capacity and the existing capacity is shut down and replaced by new capacity. One reason for this is that old generating capacity often eventually becomes subject to environmental constraints that require substantial capital investments to permit continued operation of the facility. For example, in the 2000 to 2001 period, Mirant's Bay Area generation was subject to decreases in allowable NOx emission levels that required substantial capital investments to permit continued operation of the resources. This investment appeared to economic in some of the newer units (Pittsburg 5 and 6 and Contra Costa 6 and 7 but not for the older units (Pittsburg 1-4). Pittsburg 1-4 shut down because it was more economic to invest in new capacity with lower heat rates than to make substantial investments to bring these units into environmental compliance. This same outcome should prevail in a well designed capacity market.

Similar major costs for existing capacity are on the horizon in California for units with one-through cooling systems. Treating all existing capacity as lower cost than new capacity can therefore be materially inaccurate when changes in environmental requirements take effect. Over a long period of time this shutdown of old high cost capacity might average out to between one and two percent of capacity per year. We have seen in California and Texas, however, that bursts of investment in new capacity and changes in environmental costs can cause higher rates of exit over short periods.

If this high cost existing capacity is assumed to be offered into the capacity market at a zero offer prices in determining the capacity market price but the actual offer price exceeds the clearing price in the auction, the amount of capacity procured will be less than intended as illustrated in Figure 38.[182]

Figure 38
Capacity Market Prices with High-Cost Existing Capacity



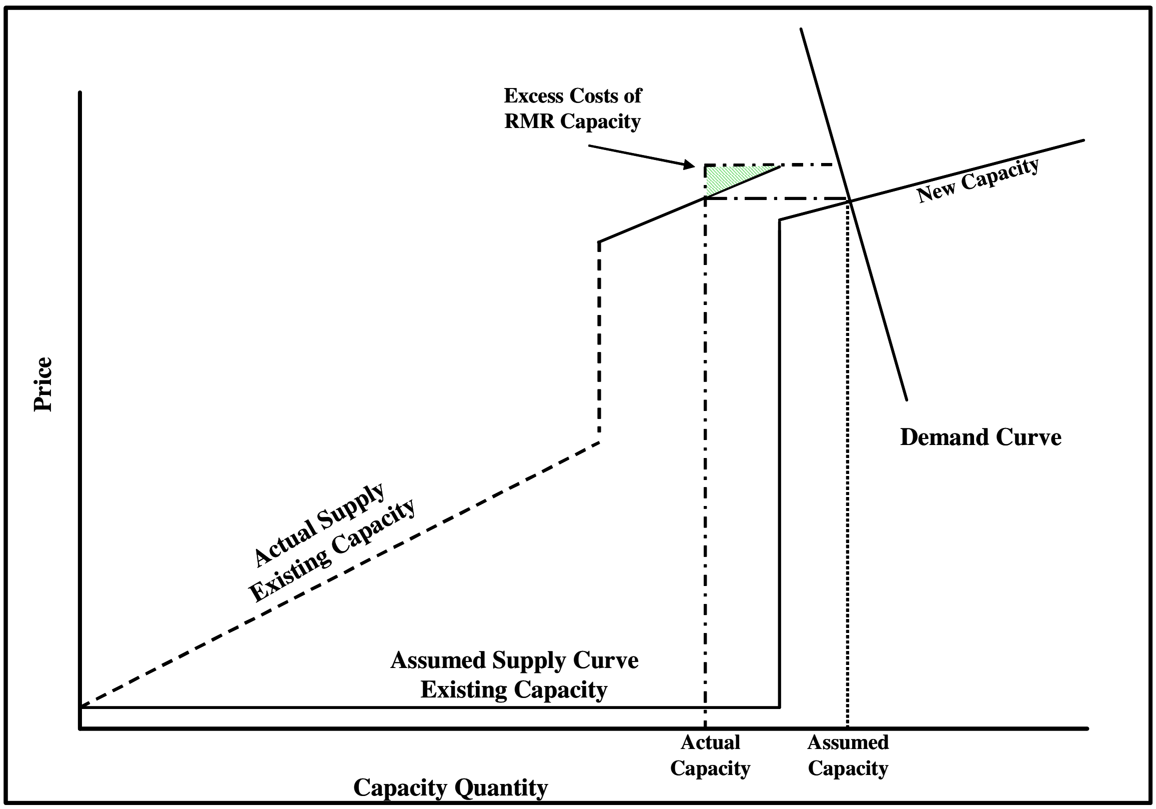
Once again, this mismatch between actual and forecasted supply would potentially have adverse reliability consequences.

---

182 These figures are simplified in that they implicitly assume that all existing capacity resources earn the same capacity and energy market revenues as the benchmark unit and differ only in terms of going-forward costs. In practice, existing units with higher energy costs and higher outage rates than the benchmark unit would earn lower revenues and might become uneconomic because of these high operating costs and outage rates rather than due to high going-forward costs.

Figure 39 shows that keeping the high cost existing capacity that did not clear in the auction available as RMR units raises the cost of meeting load relative to clearing the capacity market using the actual offer prices of existing capacity. This is an intrinsic feature of the non-market RMR process; it appears cheaper than paying market prices for capacity in the short run, but in the long run it means that lower cost entrants cannot compete with and displace the high-cost RMR unit, so load is served indefinitely by high-cost RMR units.[183]

Figure 39
Capacity Market with RMR Contracts



ISO-NE's FCM capacity market proposal in spring 2006 has a number of features that address the problematic consequences of the LICAP model assumption that all capacity is available at a zero offer price. First, the settlement agreement not only allows new capacity to set capacity prices, it also has a number of provisions that allow existing capacity to be treated as new capacity if substantial investments are needed, either to remain in compliance with environmental laws or for the units to remain economic in the

183 The costs of these RMR contracts can be spectacularly high. The RMR contract for the Pittsfield Generating Company is based on fixed O&M alone of $29,161,469 for resources with summer capacity of 150 MW. Thus, ISO-NE is paying in the vicinity of $200,000 per MW per year, far higher than the cost of capacity in New York City. See Pittsfield Generating Company, LP Cost of Service RMR Agreement, FERC Docket ER06-262-060, November 30, 2005, Attachment A and Exhibit MRK-2.

market.[184] This provision is presumably intended to address the problematic feature of the LICAP proposal which assumed that all existing capacity would be available at a zero offer price in the capacity market.

Second, the settlement agreement does not treat all existing capacity as offered at a zero capacity price. Existing capacity is permitted to submit offer prices up to .8 the estimated price of new capacity without cost justification.[185]

Third, the offer prices of capacity resources that would otherwise be permanently shutdown is in effect subject to a sliding scale offer cap at between 1.2 and 1.5 times the target price of new capacity with all of this capacity treated as if is available in clearing the market at a capacity price equal to 1.5 times the target capacity price, without regard to the actual offer price. Similarly, the offer prices of capacity resources that would otherwise be temporarily be withdrawn from the market would in effect be treated as subject to a sliding scale offer cap between .8 and 1.2 times the target price of capacity, will all of this capacity treated as offered at 1.2 times the target price in clearing the forward capacity market. Capacity requirements not covered in the initial three year out auction because of these provisions would be covered in the subsequent reconfiguration auctions.[186]

Fourth, the FCM proposal has a number of caps and floors on capacity prices. If the clearing price is above 1.4 times the clearing price, existing capacity will be paid only 1.4 times the target price while new capacity would be paid the clearing price.[187] If the clearing price calls below .6 the target price, the price will be fixed at .6 times the target price and capacity purchases prorated across resources.[188] If the total capacity offered in a zone at the starting price in the descending clock auction is less than capacity target, all existing capacity will be paid 1.1 times the target price and new capacity will be paid the starting price, with the capacity deficit made up in subsequent reconfiguration auctions.[189] If there is insufficient competition in the auction, new capacity is to be paid the clearing price and existing capacity the lower of the clearing price or 1.1 times the target price.[190]

There are a number of uncertainties as to how these offer price restrictions would operate in practice. First, the settlement agreement provides that the market monitor would review offer prices from existing capacity but it is not clear what actions would be

---

184 Settlement Agreement, Item 11, Section II B 2.

185 Settlement Agreement, Item 11, Section II.D.2.

186 Settlement Agreement, Section III.G.2.

187 Settlement Agreement, Section III.G.4.

188 Settlement Agreement, Section III.G.4.

189 Settlement Agreement, Section III.L.1.

190 Settlement Agreement, Section III.L.2.

taken if the market monitor does not agree with the capacity resource owner's evaluation of costs and risks. If the market monitor restates the offer prices for a capacity resource, what happens if the offer clears in the capacity market at the restated offer price but not at the original offer price? Would the resource be required to remain in operation at a capacity price lower than its offer price? If so, what happens if the resource owner's evaluation of the risks proves accurate and the operation of the resource is unprofitable?

Second, since the FCM market clears three years in advance of the operating year, investments in repowering and capacity expansions would presumably be contingent on the price in the auction. If the offer price of the new capacity does not clear in the capacity auction, how is the existing capacity treated?

Third, if existing capacity is offered at prices above .8 times the target price not all of the capacity requirement will be covered in the three year auction. It does not appear likely that new capacity would be available at a lower cost two years out than three years out, so if all new capacity were offered in the three year out auction, the price of capacity in the three year out auction would likely be lower than in the two year out auction. This potential outcome will presumably be reflected in the offer prices of new capacity in the three year out auction. It appears, however, that the option to take the auction price for five years is only available for capacity contracted for in the three year out auction. It would be interesting to examine the pattern of capacity prices over time and between the three-year out and reconfiguration auction produced by this structure.

Fourth, how these provisions will operate will likely depend on how accurate the target price calculation turns out be relative to the competitive market clearing price. If the target price calculation understates the cost of new capacity, the various offer price caps would more likely be binding in the three year out auction and a significant amount of capacity might be bought in the reconfiguration auction, which could raise consumer costs. Conversely, if the target price calculation overstates the competitive market clearing price, the various withholding provisions might have relatively little impact and entities possessing market power might be able to ensure that the clearing price remained close to the target price.

In summary, if some suppliers possess locational market power in the energy market, this market power is not mitigated by relying on a capacity market for resource adequacy. To the contrary, it will generally be more difficult to assess whether market power is being exercised in capacity markets than in energy markets. A critical constraint on the potential exercise of market power in power markets is the ability of load serving entities to contract forward with new entrants. The disciplining effect of entry is present both in energy and capacity markets so does not require adoption of one approach or the other to resource adequacy. Unfortunately, there appears to have been very limited contracting forward for more than a year in most restructured power markets in the U.S., including those relying on capacity markets for resource adequacy. The central problem limiting forward contracting is typically not the design of resource adequacy mechanisms but the short-term load serving obligation of most LSEs in retail access programs, but this short-term focus nevertheless limits the role of forward contracting in constraining the exercise of market power.

One way to address the potential for the exercise of market power in near term power markets given these incentive problems would be for the ISO to enter into forward contracts on behalf of those LSEs that do not contract forward on their own behalf. While this provides the benefits of competition in forward markets to consumers served under short-term contracts in retail access states, the outcome in which the ISO contracts forward for sufficient capacity to meet future load forecasts that are independent of the price of power carries other risks.

Providing for market monitor review of capacity market offer prices from existing resources is easy to suggest as a check on the exercise of market power, but this review will be less pretty when it actually has to be carried out, particularly for a capacity market system in which the net payment to suppliers depends on factors that both capacity suppliers and the market monitor are likely to have difficulty forecasting.

3. Peak Energy Market Rents

A second feature of the ISO-NE LICAP market design that was apparently motivated by market power concerns was the form of the deduction from capacity payments for peak energy market rents. All capacity market demand curve systems account in some way for the energy market profits of the marginal unit in determining the supply cost of a new capacity resource on which the demand curve is centered. NYISO bases this calculation on an estimate of the expected energy and reserve market margins. The August 2004 ISO-NE LICAP system proposed that this adjustment be backward looking/ contemporaneous, so that peak energy market rents were to be deducted from the current capacity payment received by each unit. One rationale for this approach was that calculating these margins after the fact was much more straightforward and less prone to error than trying to estimate them beforehand. A second rationale was that if actual energy market rents calculated for the hypothetical marginal resource in the capacity market were deducted from the capacity payment, this charge would eliminate the incentive of suppliers to economically withhold output in the energy market so as to raise energy prices.[191]

The various ISO-NE proposals relating to the deduction of peak energy rentals have all been vague as to how the peak energy rentals would be calculated, other than that the deduction would be calculated based on the variable costs of the benchmark unit.[192] The variable costs of the benchmark unit would include the price of gas,

---

[191] Stoft 2004, pp. 14, 19, 93-95; Cramton and Stoft 2006, p. 18. LaPlante Rebuttal, pp. 87-88.

[192] See Initial Brief of ISO-NE, pp. 47-48; Reply Brief of ISO-NE, pp. 74-80; Stoft 2004.

assuming it is a gas-fired unit, the cost of emission allowances (NOx), and variable operations and maintenance costs. The heat rate used to calculate variable costs would depend on the air temperature during the hour. Some of the complications in calculating such a deduction would be the need to take account of differences between day-ahead and real-time power prices, locational differences in power prices within the capacity zone, differences between day-ahead and intra-day gas prices, maintenance costs that vary substantially with the number of starts, not simply hours on line, non-maintenance start-up costs (fuel) and permit conditions.

There are five potentially unattractive features of the ex post peak energy rental adjustment as originally proposed in New England. First because the peak energy rental adjustment was to be calculated relative to real-time prices in the proposed ISO-NE LICAP design, the peak energy rental adjustment would have reduced the incentive of loads to participate in the day-ahead market, further compromising day-ahead unit commitment incentives. Second, there was a potential for the peak energy rental adjustment to include a material quantity of phantom rents that were not realized by real capacity resources. Of particular concern, the peak energy rental adjustment was to be calculated in the proposed ISO-NE LICAP design using real-time power prices and day-ahead gas prices. Third, the peak energy rental deduction hedged consumers against high energy prices due to shortage conditions but it would likely discourage or at least complicate long-term contracting to hedge other sources of high power prices. Fourth, the magnitude of the peak energy rental adjustment would have been very dependent on the day-ahead gas price used for the calculation. These prices are imperfectly observable and high gas prices tend to be associated with considerable dispersion in gas prices. A tremendous amount of capacity payments by consumers could hinge on the value used as the day-ahead gas price. This would have introduced a variety of risks. Fifth, if the benchmark unit had a lower heat rate and NOx allowance rate than some of the existing capacity resources, the net capacity payment received by these existing resources would vary with the level of gas prices and NOx allowance costs, introducing energy price related risks into the capacity payment stream. Each of these issues is discussed further below.

Consider the first issue, the implications of calculating the peak energy rental adjustment relative to real-time power prices. An ongoing peak energy rental deduction from the capacity price based on real-time power prices would affect incentives for participation in the day-ahead market. It was observed above that capacity market resource adequacy systems inevitably somewhat reduce the efficiency and reliability of the day-ahead market and unit commitment process by capping real-time prices. This reduces the financial consequences to LSEs of understating their real-time load, particularly on days when real-time load is uncertain and could be materially higher or lower than its expected value. If an ongoing peak energy rental deduction is calculated based on real-time, rather than day-ahead power prices, this deduction may tend to further undermine the effectiveness of the day-ahead market. With an ongoing peak energy deduction calculated in this manner, LSEs would know that to the extent real-time prices rose above the benchmark price used in calculating the peak energy rental, the energy payments in excess of the benchmark price would be credited back to the LSEs through the capacity payment deduction.

To the extent that this behavior lead to a material difference between day-ahead and real-time prices, one would expect suppliers to attempt to raise their offer prices in the day-ahead market to reflect expected real-time prices, however, this might be prevented by the application of market power mitigation rules to supplier offer prices. Alternatively, virtual demand bidders could bid up day-ahead prices, expecting to realize profits in selling the power at higher real-time prices. Similarly, small LSEs whose activities would not impact market prices would likely continue to participate in day-ahead markets.

While there would continue to be arbitrage incentives for virtual traders to drive day-ahead and real-time prices together through virtual demand bids in a market including such a retrospective adjustment for peak energy market rentals calculated based on real-time prices, the introduction of such an adjustment would impact the balance of incentives deterring intentional underscheduling of load in the day-ahead market. The supply elasticity of virtual demand bids is important in deterring intentional underscheduling of load in part because virtual demand purchases increase the proportion of the underscheduling entity's load that must be purchased at real-time prices, relative to the reduction in total day-ahead load. While the elasticity of virtual demand offers is important in deterring such underscheduling, part of the deterrent is the cost of purchasing the underscheduled load at real-time prices that might be considerably higher than actual day-ahead prices and higher than the cost of meeting this load had it been scheduled in the day-ahead market. By reducing the penalty to underscheduling, a real-time based adjustment for peak energy rentals may materially disturb the current balance in Northeastern day-ahead and real-time energy markets.

It is unambiguous that such an ongoing deduction for real-time peak hour rents would at the margin shift the financial impact of high real-time prices resulting from underbid load from the shoulders of the LSEs that underbid their load and onto suppliers and also would tend to push both suppliers and LSEs into the real-time market. This feature of the peak energy rental deduction would exacerbate the incentive problems arising from capacity market with artificially low prices during reserve shortage conditions, as discussed in Section III.E.6 above. The uncertainty is whether the incremental effect would be large or small. Such changes would tend to depress day-ahead and load prices, reducing the unit commitment in the financial market and pushing more of the unit commitment into the reliability commitment, adversely affecting overall reliability to the extent that the reliability commitment leaves less room for error than would be the case if LSEs bore the consequences of high real-time prices. This outcome would be particularly undesirable during winter conditions in combination with the kind of capacity resource performance incentives envisioned as part of the ISO-NE LICAP or FCM systems. If underbidding by LSEs caused the day-ahead generation schedules to provide a poor indication of the real-time dispatch, powerful capacity market performance incentives could make gas-fired capacity market resources reluctant to sell gas in the day-ahead gas market because of uncertainty regarding their own real-time gas requirements. While this might be good for electric system reliability, it could artificially inflate winter gas market prices and adversely impact gas pipeline operation.

A second issue in applying an ex post peak energy rental deduction to capacity payments is the potential for misstated peak-energy rentals arising from the comparison of day-ahead gas prices and real-time power prices. The peak energy rental deduction included in the ISO-NE August 2004 LICAP proposal was apparently to be calculated based on real-time energy prices and day-ahead gas price.[193] Under the relaxed gas pipeline balancing rules that are in effect during the summer, gas prices are typically not very volatile between the day-ahead and intra-day markets, so the peak energy rental would generally be little different in the summer whether calculated based on day-ahead or intra-day gas prices. This is not the case, however, during the winter when gas system reliability requires gas pipelines and distribution companies to enforce gas balancing rules that tightly constrain in day gas system withdrawals based on day-ahead gas schedules. Colder than anticipated weather, among other things, can cause intra-day gas prices to substantially exceed day-ahead gas prices for power producers seeking to purchase gas in the intra-day market. These price spikes in the intra-day gas market have been a feature of most of the winter power system reliability crises. When intra-day gas prices substantially exceed day-ahead gas prices, real-time power prices typically are much higher as well, reflecting both the higher value of the gas used to generate power and often an accompanying increase in cold weather related power demand. If day-ahead prices gas prices were moderately high with corresponding moderately high power prices but intra-day gas prices are extremely high with corresponding extremely high power prices, the peak energy rental deduction will be quite large if it is calculated based on day-ahead gas prices and real-time power prices. In practice, however, most of the power market suppliers that scheduled gas at day-ahead prices would have also sold their power in the day-ahead market, while power market suppliers whose generation was not scheduled to operate in the day-ahead market but find themselves operating in real-time would be selling power at high real-time prices but likely also buying gas at spectacularly high intra-day gas prices.

The seemingly simple solution of calculating peak energy rentals based on real-time energy prices and intra-day gas prices would not be practical because intra-day gas

---

193 The ISO-NE LICAP proposal was extremely vague on how the peak energy rental deduction would actually be calculated. It is understood from various presentations and discussions that the day-ahead gas price was to be used in calculating these rentals. Stoft 2004, p. 100, refers to the "daily gas price" for the calculation of critical hours and Cramton and Stoft 2006, simply refer to variable cost without specifying the gas price, NOx allowance cost, heat rate or run-time to be used in calculating the variable cost.

markets are generally very thin. It might seem that these complexities would be avoided if the benchmark unit used to calculate the peak energy rental were a dual-fueled unit as prices for #2 and #6 oil are observable at many distribution points and can be used to value the replacement cost of fuel stocks. During some winter reliability crises (Texas 2003, PJM 1994), however, ice-covered roads have prevented resupply so calculation of peak energy rental deductions based on unlimited fuel supplies at the distribution point price would have misstated the actual peak energy rentals of real units.

Some of these limitations of an ongoing deduction for peak hour rentals could be avoided by shifting the application of the deduction to the day-ahead market. If peak hour rentals were calculated based on day-ahead market prices, this would tend to push power consumers into the day-ahead market, rather than pulling them out of the day-ahead market into real-time. Similarly, since day-ahead gas prices are generally observable, there would be much less error in the calculation of peak energy rents during winter conditions if the calculation were based on day-ahead gas prices. While the owners of quickstart resources might earn additional profits from operating during real-time shortages, the expected value of these earnings would be reflected in capacity offer prices. A limitation of this approach is that without full scarcity pricing in real-time, LSEs still would not have fully efficient incentives to bid their load into the day-ahead market but at least a peak energy rental deduction based on day-ahead power and gas prices would not make day-ahead scheduling incentives worse than they already are.

Another potential for phantom peak energy rental deductions would arise if NOx allowance prices rose and these costs were not reflected in the peak energy rental deduction or if the hypothetical benchmark unit had a very low emission rate relative to existing units. Although combustion turbines are quick-starting units, they have a start-up process in which they burn gas without generating power. In addition, maintenance costs depend on both run-time and starts so short run times dramatically increase variable maintenance costs. These costs need to be accounted for in the PER deduction. Other phantom deductions could arise from the use of understated heat rates and excessive starts. The actual heat rate performance of combustion turbines depends on ambient air temperature; high temperatures reduce conversion. Unless the PER deduction is calculated using actual temperature adjusted heat rates or worst case values, there would be a phantom per peak energy rental on hot summer days. This potential for phantom deductions in peak energy rentals would be a factor in the level of capacity prices, and is a reason why the actual capacity supply curve might be higher than estimated by the ISO.[194]

ISO-NE proposed to address the potential for such phantom peak energy rental deductions through a “bias” factor that would adjust the calculated peak energy rental

194 These kinds of issues were pointed out by Ken Bekman, among others; see Ken Bekman, Docket ER 03-563-030, Exhibit CEE-1, p. 14.

based on the revenues of actual units.[195] The details of how this bias factor would be calculated were deferred for development later.

A third issue concerning a peak-energy rental adjustment is its impact on LSE hedging incentives. Some of the discussion of the rationale for an after the fact peak-energy rental adjustment refers to the deduction hedging consumers against increases in real-time prices,[196] but even if the deduction operates as intended it fully hedges consumers only against real-time gas price spikes and real-time capacity shortages. It does not hedge consumers against increases in the general level of gas prices nor increases in other costs such as emission allowance costs for the benchmark generator. Unless the characteristics of the benchmark generator were very different from the characteristics of actual generators, virtually none of the high power prices that occurred in California from August 2000 through June 2001 would have been offset by a peak energy rental deduction as the high power prices reflected extremely high gas and NOx allowance prices.

Similarly, none of the gas price driven increase in power prices for a benchmark generator during 2005 and 2006 would have resulted in increased peak energy rental deductions so consumers would not have been hedged against the increases in the price of power. While the peak energy rental deduction fails to hedge consumers against power price increases attributable to changes in gas and emission allowance costs, it appears likely to deter other power price hedges.

While a peak energy rental might deter capacity resources from exercising market power to raise prices above the heat rate of the benchmark unit, power prices might still exceed the bench mark level if set by demand response, imports, reserve shortage pricing or the costs of high heat rate or high emission rate existing capacity. Capacity resources would not be willing to enter into a conventional price hedging contract with customers, either in the form of a contract for differences or in the form of a call contract, because the peak energy rental deduction would in effect have already sold a portion of their energy market revenue stream forward.

Suppose, for example, that a resource had entered into a forward call contract with a strike price of $70/MWh against the price of power.[197] The generator would earn $70/MWh when called, but would have to pay buyer the difference between the actual price of power and $70/MWh when called. In addition, suppose, that an ex post deduction of peak energy rentals were also applied to the capacity payment and the variable cost of the benchmark unit was $100/MWh. Then anytime the power price rose above $100/MWh, the seller would have to forgo revenues equal to twice its revenues in

---

195 Mark Karl, Exhibit ISO-39, p. 38-39.

196 Cramton and Stoft 2006, pp. 5, 12.

197 To keep the example simple, we will ignore the difference between day-ahead and real-time prices.

the energy market. It is possible that the forward power market in the capacity market region could adapt to such an ex post peak energy rental deduction by developing an option contract that hedged loads only for variations between a fixed strike price and the gas and NOx allowance varying benchmark price, but this is uncertain.

A fourth significant feature of an after the fact peak energy rental deduction is that the index price used to measure gas prices and allowance prices will be extremely important. It should be kept in mind that the indexes of day-ahead gas prices are a sample of the market transactions at a particular location that take place over a period of time. The reported day-ahead gas price is not a single market clearing price but is actually the volume weighted mean of the reported prices. High priced days in the gas market are often volatile days and transaction prices may move substantially over the trading period, so the average price is not necessarily the price at the end of the trading period. On a day in which the gas price rises over the trading day, increased transaction volume at the market price at the end of the trading day would raise the average gas price for the day and reduce the peak energy rental. Similarly, on a tight supply day the volume of transactions might be relatively low and the average price could be materially impacted by a small number of high or low priced transactions. Another interesting feature of the ISO-NE peak energy rental adjustment will be to observe its interaction with the gas market and whether the gas price index becomes a major source of regulatory risk.

A fifth significant feature of an after the fact peak energy rental deduction based on the estimated variable costs of a hypothetical unit is that the net capacity payment of resources with higher heat rates or higher NOx emission rates than the benchmark unit will vary with the gas price and NOx allowance price. This will provide consumers with a partial hedge against increases in gas prices and NOx allowance costs as consumers would only bear the gas and allowance price risk of serving load with the benchmark unit. If the bench market unit is a modern unit with a lower heat rate and lower emission rate than existing units, the capacity payment received by existing capacity may be very risky, varying with gas and NOx allowance prices. While the outcome that resources with a higher costs than the benchmark unit would need to have expected lower going forward costs in order to compete in the capacity market is consistent with the operation of an efficiently structured energy only market, the ex post deduction of peak energy rentals based on the cost of a hypothetical unit has effects that are quite different than in an energy only market.

If gas prices rose dramatically during the year in an energy only unit, power prices would rise commensurately and while high heat rate units might find their operation constrained by higher heat rate oil and coal resources, the high heat rate gas units would be able to still recover their going forward costs in energy and reserve prices during shortage conditions. With a peak energy rental adjustment, however, increases in gas and power prices could result in a peak energy rental adjustment that effectively eliminates the capacity payment to high heat rate resources, providing little incentive for the resource owner to keep the resource available even though it is needed for reliability.[198]

The owner of a high heat rate resource could in principle hedge itself against reductions in the capacity payment resulting from increases in gas or allowance prices by purchasing gas and NOx allowances forward, so that it would earn profits on the gas and NOx allowance position to offset the reduction in the capacity payment. The peak energy rental adjustment would be difficult to hedge, however, because the magnitude of the peak rental adjustment would depend on the heat rate the marginal unit not the high heat rate capacity resource so the hedge would need to be based on the hypothetical gas consumption of the benchmark unit, not the actual gas needs of the high heat rate capacity resource.

This fifth issue can be largely addressed by tying the peak energy rental deduction to the characteristics of existing high heat rate, high emission rate units, which would greatly reduce the variability of the capacity payment with gas and NOx allowance prices. Another facet of this issue is that the peak energy rentals in the ISO-NE LICAP proposal were apparently to be calculated without regard to energy limitations, but in practice the benchmark unit would be subject to permit conditions limiting the hours of operation in many zones. This would likely also be the case in areas outside New England. If peak energy rentals are calculated without regard to such operating hour limitations, it is likely that the target capacity payment would be insufficient to support even the benchmark unit, because the benchmark unit would not be able to operate for the number of hours assumed in calculating peak-energy rentals.

The retrospective adjustment for peak hour rentals in the August 2004 LICAP design evolved over time into a prospective deduction.[199] With a monthly capacity auction, monthly variations in the energy rental deduction would be reflected in the capacity offer prices of both new and existing capacity resources. Future deductions for past peak hour rentals that exceed the updated level of future energy market margins would simply raise offer prices for future capacity and need to be taken into account in estimating the capacity supply curve.[200] To the extent that the deductions for abnormal past peak energy rentals are spread over the remainder of the current capacity market year, abnormally high deductions for past market revenues will reduce the availability

---

198 This outcome could be avoided by allowing the capacity payment to go negative, but that might make the net capacity payment so risky that capacity offer prices would increase substantially.

199 Settlement Agreement, Item 11, Section V.B.2. Initial Brief of ISO-NE, pp. 47-48. LaPlante Rebuttal, pp. 14-15, 84-86.

200 Under the ISO-NE's proposed "market power mitigation" mechanism, the capacity price would be determined using an assumed zero offer price for capacity. If the determination of the demand curve ignored differences between the past level of peak energy rentals and expected future levels (which appears to have been the intent), this would tend to magnify the adverse reliability impact of the market power mitigation mechanism.

incentive during those remaining months. This will not be important if the costs of making the capacity available for the remainder of the year are sunk by the time the peak energy rent deductions reduce capacity payments but it will be important if the capacity payment incentive is needed to ensure that resources are manned to start on short-notice, that gas is scheduled during the winter, etc. Thus, if unusually large peak energy rentals during July and August reduce the capacity payment during January and February to very low levels, the availability incentive under the capacity market system would be greatly reduced leading to reliability events that could have been avoided under an energy market pricing system.

The ISO-NE FCM proposal also includes a peak energy rental adjustment but it differs from the kind described above in a number of important respects. First, instead of the peak energy rental being calculated relative to the characteristics of a benchmark unit corresponding to the kind of unit that would be built to meet incremental capacity needs, it would be calculated based on a proxy for existing high cost units. Thus, the proxy unit would have a 22,000 heat rate.[201] It is not explained whether peak energy rentals would still be calculated using real-time energy prices and day-ahead gas prices or in some other manner.[202] Although not specified in the FCM settlement, it would presumably be assumed to have correspondingly high NOx emission rates and variable operating and maintenance costs.[203] As noted above, the deduction of peak energy rentals would also be partially prospective. The substitution of the assumed higher heat rate would lessen the adverse impact of the peak energy rental deduction on LSE incentives to participate in the day-ahead market (item 1 above) and would eliminate the risks for existing units from changes in gas and NOx allowance prices (item 5 above).[204]

Another important feature of the FCM proposal is that LSEs that self-provided capacity in the auction would be exempt from the peak energy rental deduction. Self-provided capacity in the auction not only covers the capacity of vertically integrated load serving entities but also includes capacity contracted for the load serving utility prior to the auction. These provisions would not only enable vertically integrated utilities, such municipals in New England, to avoid the risks created by the peak energy rental deduction, but would also allow LSEs to contract forward on their own for capacity and energy and cover their capacity requirements outside of the forward auction and also in effect opt out of the peak energy rental deduction, facilitating forward contracts for energy as well as capacity. To the extent, however, that LSEs find it desirable to hedge part of their energy costs closer in time to the operating year than three years forward, for

---

201 Settlement Agreement, Item 11, Section V.B.1.d.iii.

202 Settlement Agreement, Item 11, SectionV.B.1.

203 Settlement Agreement, Item 11, Section V.B.1.

204 Assuming that the calculation of the peak energy rental deduction would also reflect the NOx allowance costs of the existing marginal generation.

the kind of reasons discussed in Section G above, the existence of the peak energy rental deduction may somewhat retard forward contracting even under the FCM system with the option for opting out of the peak energy rental deduction through self-provision of capacity in the auction. If such a peak energy rental deduction is eventually implemented in New England, it will be interesting to examine its impact on the forward bilateral contract market.

4. Addressing Market Power

Underlying the discussion of market power in capacity markets is a more general question of how RTOs and state and federal regulators should address the potential for the exercise of unilateral market power in electricity markets. Consider first the case of locational market power, i.e., generation resources that possess locational market power when particular transmission constraints are binding. At one extreme there are generators that possess market power, the ability to set prices above the competitive level by economically or physically withholding output, during a few hours of the year when particular patterns of transmission and generation outages are present, but that otherwise lack market power and whose operation is economic on a going forward basis independent of the revenues in these hours. Imposing market power mitigation rules in the energy market that constrain large changes in the offer prices of such resources during these unusual conditions is a workable approach to market power mitigation that does not require further intervention in energy markets.

Alternatively, however, suppose that this resource has such high costs that it only operates when this constraint is binding and therefore its ability to recover its going forward costs depends entirely on its offer prices during the hours in which this constraint is binding. In this situation capping the offer prices of the resource resolves the market power problem in the short-term, but the offer price caps lead to further problems because some mechanism is required for the resource to recover its going forward costs. Moreover, unless there is some mechanism for defining the market price of power when this constraint is binding, there is no potential for competition from entrants, because LSEs would not offer high enough contract prices to support an entrant if spot market prices were low. Shifting the market power problem into the capacity market solves nothing in this situation as if the supplier has market power in the energy market, it has market power in the capacity market as well.

FERC and the relevant state regulators need to pose an initial question in these circumstances of whether it makes economic sense to divest this resource from the distribution company in the first place or whether it should be retained in the rate base of the distribution company. If there were large efficiency gains from shifting the management of the resource out of the rate base of the distribution company, the market power of the resource could be addressed on a prospective basis through a divestiture contract that provided a call option on capacity at a predefined price (indexed) in exchange for a pre-defined (indexed) payment of going forward costs. It would then be the responsibility of the distribution company to periodically extend this contract sufficiently prior to its expiration that the resource must compete with transmission upgrades or new generation projects. It is important to recognize that adding a capacity market to the energy market does nothing to solve the market power problem.

Next consider the case of resources that do not possess locational market power but could occasionally possess market power in the energy market during high load conditions. This is the pivotal supplier problem. Suppose that load is 27,000MW, required reserves are 1,800 MW and there is 28,830 MW of capacity available either on line or in quick start units that can provide reserves. There is no shortage of capacity, but the owner of a 500 MW unit with a 2 MW/minute ramp rate could drive prices to the offer bid cap by offering 100 MW of the unit's capacity at the bid cap. 60 MW of this high-priced capacity could be used to provide reserves and 30 MW would be excess, but the RTO would need to dispatch 10 MW of the high offer price capacity in order to maintain reserves and meet load. If the RTO had effective shortage pricing that would set energy and reserve prices at high levels during true shortage conditions without regard to bids, then this potential for the exercise of market power during shortage conditions could be addressed through relatively loose offer price caps that prevented major changes in offer prices during shortage conditions.[205] There is an analogous pivotal supplier problem in near-term capacity markets, but mitigation is more complex than in energy markets as it requires analyzing the going forward costs of each market participant in these short-term capacity markets, which are poorly defined. In a longer term forward capacity market in which entrants can compete with incumbent suppliers there is less potential for a pivotal supplier problem unless individual suppliers have quite large market shares. Overall, capacity markets do not contribute to addressing market power problems.

I. Conclusions

Five conclusions can be drawn from the discussion above. First, capacity market systems need to be based on locational capacity markets. Experience has shown that a general deliverability requirement will not provide the incentives necessary to maintain reliability. The NYISO locational capacity market has worked reasonably well over, at least from a locational incentive standpoint, over the five years it has been in operation but it has not yet been seen if this successful experience can be generalized or even if the New York design will avoid locational problems in the longer term. While locational

[205] This incentive to exercise market power would not exist if the supplier had sold its capacity forward, but absent effective market power mitigation, a rational supplier would not sell capacity forward in this situation except at a price reflecting its potential profits from the exercise market power, see Scott M. Harvey and William W. Hogan, "California Electricity Prices and Forward Market Hedging, October 17, 2000.

capacity market system provide appropriate incentives for capacity to locate within the capacity zones, zonal definitions may not reflect all of the relevant transmission constraints.

Second, greater diversity in the resources providing capacity increases the importance of providing appropriate incentives for capacity resources to be available during shortage conditions. While the current UCAP systems are suitable for taking account of forced outage risk across thermal units, experience has shown the need for improved performance incentives for gas fired generation, energy limited resources, and resources with limited availability such as intermittent resources. Performance incentives tied to availability during reserve shortage conditions have the potential to provide more efficient availability incentives than UCAP systems but cannot replicate the incentives for generation and load provided by an energy-only market with effective shortage pricing.

The limitations of a capacity market system are particularly acute in providing efficient incentives to energy consumers. Capacity market systems do not provide fully efficient incentives for LSEs to participate in the day-ahead market and provide particularly poor incentives for real-time demand response. Capacity market designs that provide for forward market contracting for capacity by the ISO have the potential for creating stranded costs by contracting for capacity to meet load that would not exist at the prices required to support that capacity. This potential is particularly acute in the west where resource adequacy must take account of the hydro cycle.

Third, it is likely that the greatest potential advantage of a capacity market resource adequacy mechanism relative to energy only markets is in reducing the regulatory risk associated with energy only markets in a retail access environment in which few LSEs enter into multi-year forward contracts. A difficulty in achieving these benefits is that the same incentives that deter LSEs from entering into multi-year forward contracts to hedge energy and reserve costs also deter LSEs from entering into multi-year forward capacity market contracts. Proposals to address this problem by having the ISO coordinate forward capacity purchases on behalf of retail access load that would otherwise be unhedged have been described but not yet implemented, so the success of this approach will need to be evaluated once the details have been developed and a system implemented over a period of years. An important disadvantage of this approach is that the ISO is effect required to purchase capacity forward to meet forecast load without regard to the price of that power. This course of action carries the risk pointed out above of giving rise to substantial stranded costs if capacity is purchased for price sensitive load that would not consume power at the implied price of energy and capacity. A fundamental policy decision is whether retail access incentive problems should be addressed through ISO capacity and energy market design or through regulatory reform of the retail access design.

Fourth, while capacity market payments are sometimes used as the basis for mitigation of energy offer prices to address the exercise of market power in short-term markets, the introduction of capacity markets does not mitigate market power. Moreover, there is no actual experience to date showing that it is easier to identify the exercise of market power in capacity markets than in energy markets, and the reverse is much more likely. The potential for the exercise of market power can often be addressed (absent material barriers to entry) through forward contracting with entrants by consumers, but the incentive to enter into such contracts for either capacity or energy has been undermined by retail access programs. While forward capacity markets reduce the potential for the exercise of market power, if all capacity is purchased at a fixed point in time in forward markets, the threat of entry may not be fully effective in constraining the exercise of market power by incumbent suppliers. The long-run capacity market competitive offer price level for existing capacity is not zero, and can at times be quite high, so constraining existing capacity to be offered at a zero price will likely cause the capacity market to collapse into a system based on RMR contracts. The ability of regulators (including the RTO market monitor) to distinguish between the exercise of market power and competitive behavior by existing capacity is critical to capacity market performance and there is little reason to expect good outcomes from such a process.

Fifth, there is a potential for both capacity and energy-only market designs to fail to provide the incentives required to maintain the intended level of reliability. While a capacity market system may appear able to assure the intended level of reliability by specifying the capacity level required to maintain reliability, maintaining reliability is more complex that simply specifying a nominal or even expected capacity level. Maintaining conventional levels of reliability at acceptable cost requires delivery of a load shape of energy and reserves over the year, to the location of load. It needs to be kept in mind that the conventional Monte Carlo approach to simulating capacity market requirements starts with the characteristics and location (including its location within a zone) of existing units and derives the required level of capacity. This non-sequitur has worked historically in part because the resources that would be used to meet load were in practice largely fixed by the time the capacity market was evaluated so that variations in the actual mix of resources acquired in the capacity market did not affect the reliability analysis. This will not be the case if the capacity market is moved forward in time and the unit characteristics modeled in the reliability analysis potentially differ from the characteristics and locations of the resources cleared in the capacity market.

IV. CALL CONTRACT CAPACITY DESIGN PROPOSALS

Subsequent to the initial implementation of capacity markets in PJM and NYISO in 1998 and 1999, several proposals have been outlined for using call contracts both to hedge consumers against volatile energy prices and as a resource adequacy mechanism. The general structure of the proposals is that load serving entities would be required to enter into call contracts hedging a specified portion of their loads, much like load serving entities are required to purchase capacity credits under a capacity market system.

These mandatory call contracts would typically have several elements.[206] First, the call contract would have a strike price that is less than the price cap in market, so that suppliers that fail to cover their call contracts would incur a financial cost. Second, in addition to the market based penalty for non-delivery, there would be an additional penalty for generation that does not deliver power in a shortage. Third, the call contracts would be for power delivered to load (i.e., they would be locational and the power would be sold on a delivered basis). Fourth, the call contracts would be entered into several years forward, allowing time for construction of new capacity. Fifth, the call contracts would be required to be backed by physical capacity. Sixth the call contracts would cover at least one year.[207] Although not specified in most of the proposals, the call contract should settle in the day-ahead market, putting the load forecasting burden on the LSE, and the contract should cover load plus an appropriately measured share of reserves and regulation.

The call contract approach could also encompass demand response, which could be treated as a providing call contract coverage by reducing load.

A critical feature of these systems is that the spot price of energy would not reflect the actual value of power during reserve shortage conditions. The effectiveness of the resource adequacy role of these contracts therefore depends on the second element, the administrative penalty for non-delivery during shortage conditions. If the non-delivery penalties were sufficiently onerous suppliers would not be willing to enter into these contracts unless the supplier was hedged with a resource or contract to cover the contract and avoid the non-delivery penalties. The price of call contracts would then have to be high enough to cover the cost of physical resources to meet load. As with capacity market systems, it would therefore be critical to the effectiveness of call contracts in sustaining resource adequacy that the administrative penalties for failure to deliver or failure to contract be sufficiently high to induce loads to contract and suppliers to perform.[208]

---

206 Miles Bidwell, "Reliability Options: A Market-Oriented Approach to Long-Term Adequacy," Electricity Journal, June 2005, pp. 11-25; Oren, Shmuel, "Generation Adequacy via Call Options Obligations: Safe Passage to the Promised Land," September 2005; Singh, Harry; PG&E National Energy Group, "Call Options for Energy: A Market Based Alternative to ICAP," October 16, 2000.Vazquez, Carlos, Michel Rivier, and Ignacio Perez-Arriaga, "A Market Approach to Long-Term Security of Supply," March 2001.

207 Some approaches have a multi-year tranche or allow entrants to convert into a multi-year obligation.

208 Cramton and Stoft assert that call contracts intrinsically fail to solve the missing money problem, i.e., will fail to sustain the construction and operation of enough capacity to maintain reliability but do not explain why this would be the case, Cramton and Stoft 2006, pp. 36-38.

Two advantages of such a call contract system relative to a capacity market system such as the ISO-NE LICAP system are that the call contract approach potentially provides a better link between the capabilities and obligations of individual units and that it supports rather than undermines forward hedging of energy costs. The first advantage is that call contracts can be defined for a fixed number of hours per year, month or day, consistent with the characteristics of the resource.[209] This would be particularly advantageous for making efficient use of limited energy resources, as the resources would be able to make a specific supply commitment consistent with the characteristics of the resource. It would be the responsibility of the LSE to enter into a portfolio of call contracts fitting the load shape. Under a capacity market with performance penalties on the other hand, limited energy resources would have an unlimited obligation to perform and could fail to earn any capacity payments as a result of sustained shortages attributable to the outage of baseload units.

A second advantage of a call contract resource adequacy mechanism would be that it would provide a hedge of energy market costs (whether due to shortages, changes in gas costs, or changes in NOx allowance costs) and its form would be consistent with existing trading conventions.

A critical problem with these proposals is in the application of the administrative penalties for non-performance during reserve shortage hours in which the spot energy price is low. If the spot energy price is low during shortage conditions, the administrative penalties for non-performance must be structured as some kind of availability test for physical generating capacity, just as under the LICAP or FCM proposals. The penalties and performance incentives therefore actually have nothing to do with the call contract, they are simply penalties for capacity that fails to perform during the shortage hours.

This underlying reliance on a capacity based performance metric to define call contract performance obligations also leads to problems in measuring deliverability. If energy prices do not rise to shortage levels during shortage conditions within a load pocket, then there is no penalty to not being financially hedged with FTRs so the fact that the power is sold on a delivered basis is meaningless. Even if resources supporting call contracts were required to purchase FTRs from generation sources to load, the source of the FTR need not match the day-ahead or real-time generation source, and the actual resources available in the day-ahead market may not be deliverable.

A third feature of these proposals is that they do not directly address market power. At what price must capacity be offered in the market? This is fine if in fact there is no supplier that possesses market power within a load pocket within the four year

---

[209] This is a potential advantage of a call contract approach but it is not present in all call contract type proposals, some of which would apparently impose an 8760 hour call contract obligation on all resources which would have much the same disadvantages as the availability based performance penalties based on delivering a load share of capacity over 8,760 hours in the ISO-NE LICAP proposal.

window for forward contracting but it is a problem if there is a potential for withholding within some load pockets in which there asymmetries between the cost of adding capacity at existing and new sites.

A fourth feature of these proposals that needs to be carefully considered is the level and timing of the required hedge. It may not be commercially reasonable to fully lock in the cost of power from all generation five or six years in advance, as the suppliers may not be able to hedge their own energy market supply risks that far out. It may be in the interest of loads to lock in a portion of the baseload energy supply cost five or six years in advance but to defer hedging the energy price for another portion of consumer load until closer to the operating year. In principle, this could be addressed by requiring a certain proportion of the load be hedged at various dates, perhaps with different rules for hedging capacity and energy costs.

It is not a foregone conclusion that all load should be hedged against variations in power prices. Indeed, in practice the least cost approach to managing power system reliability over the Western Hydro cycle, for example, may be to not hedge the power costs for price sensitive industrial demand, raising power prices during low hydro years to the point that these customers reduce their load. The alternative of building six thousand or so megawatts of capacity that runs once every six or seven years during a low hydro year could be spectacularly expensive.

The fifth question is who signs these contracts for load in retail access states in which no one has an obligation so serve load five years from now. If the current LSE is required to enter into a contract to serve load five years from now at prices that may turn out to be dramatically out of the money five years from now, how do we ensure the supplier against default? Conversely, how do we ensure the LSE against default by the supplier if the power price turns out to be spectacularly in the money five years from now.

Some of these problems could be address by modifying the call contract approach as suggested by William Hogan.[210] Instead of requiring call contracts and imposing penalties for failure to perform, the ISO could implement shortage pricing in real-time and day-ahead markets and require LSEs to enter into call contracts hedging a portion of their load, to address the political problem, ensuring that all loads are hedged to a considerable degree against variations in energy prices. This approach would avoid the complications discussed under items one and two above. In addition, market power could be addressed by monitoring real-time behavior, since forward prices would be driven by real-time shortage prices. This approach would also allow LSEs more

210 See William W. Hogan, "On an 'Energy-Only' Electricity Market Design for Resource Adequacy," September 23, 2005.

flexibility in the timing of their hedging, as long as they complied with a minimum standard for forward hedging.

V. HYBRID RESOURCE ADEQUACY DESIGNS

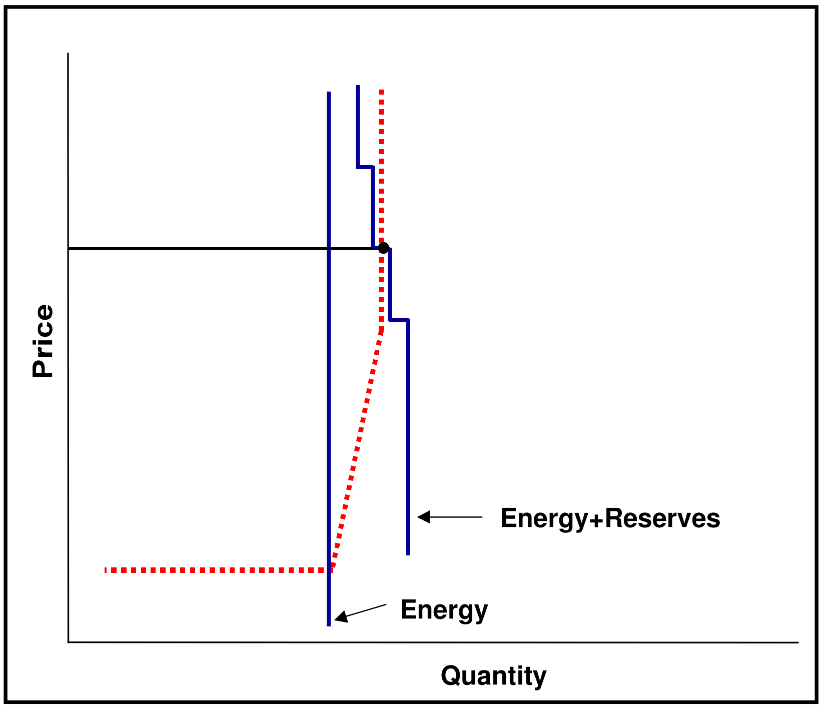
Most of the discussion above has focused on the polar alternatives – capacity market systems under which resources derive little margin from prices during shortage conditions or energy-only markets in which resources derive all of their going-forward costs from energy market margins during shortage conditions.

In view of the very difficult performance incentive problems characteristic of capacity market systems and the high regulatory risk associated with energy-only market designs, a potential short-run alternative might be to trend toward a hybrid system in which a large enough proportion of capacity resources are going forward costs are recovered in the capacity market to reduce regulatory risks while a large enough proportion of those going-forward costs are recovered in energy markets to induce reasonably efficient performance. The NYISO market most closely approximates such a hybrid and is trending further in this direction. Several elements of the NYISO market design are relevant.

First, the explicit reserve markets of the NYISO provide an additional relatively stable income stream for the marginal capacity resource, which should be a 10-minute combustion turbine located east of Central East that provides reserves, even if it is not dispatched for energy. The historical expected reserve market earnings of around $10,000/MWyear fall far short of what is required to keep the marginal unit in operation but provide incentives for availability and deliverability as these revenues are not received if the resource is not available or not dispatchable.

Second, the reserve demand curve implemented for 30-minute reserves by the NYISO prior to the summer of 2002 somewhat raises energy prices during shortage conditions, even if suppliers bid their costs. Moreover, it addresses the potential for the exercise of market power through economic or physical withholding during those conditions by in effect making the residual demand curve facing a supplier with market power in the energy market more price elastic than would otherwise be the case, as shown in Figure 40.[211]

Figure 40
NYISO Reserve Demand Curve



Third, the reserve shortage pricing introduced for 10-minute and spinning reserves on February 1, 2005 allows real-time energy and reserve prices to reach several thousand dollars per MW, as a result of reserve shortages, even with the $1,000/MWh bid cap. This shortage pricing system provides marginal incentives for generator performance during strained system conditions although the shortage values are currently set far too low to obviate the need for capacity payments to the marginal generator. Current reserve shortage values will produce margins approaching $2,000 only during conditions that are planned to occur for only a few hours a year. Nevertheless, there were a number of

---

211 The reserve demand curve reduces the quantity of 30-minute reserves based on the shadow price of 30-minute reserves. If the shadow price of 30-minute reserves reaches $50/MW the amount scheduled can fall by up to 200 MW to 1,400 MW. If the shadow price of 30-minute reserves reaches $100/MW, the amount scheduled and fall by up to 400 MW to 1,200 MW. If the shadow price of 30-minute reserves reaches $200/MW, the amount scheduled can fall by up to 600 MW to 1,200 MW.

hours with quite high real-time reserve prices during Summer 2006, generating shortage revenues. The importance of these energy market revenues could be gradually increased over time through increases in the shortage costs used to determine prices.

One can think of the NYISO shortage pricing rules as one way of addressing this potential incentive problem by attempting to ensure that the marginal capacity market resource recovers a meaningful proportion of its going-forward costs from the energy market during shortage conditions. Units whose 350 hours of forced outage occur during reserve shortage hours around the summer peak could forgo much more than the outage cost under the UCAP system.

One can therefore think of the evolution of NYISO reserve markets, shortage pricing and reserve demand curve to date not as obviating the need for a capacity market but as tending to ensure that the marginal generator recovers an appreciable proportion of its going-forward costs in energy and reserve markets and is thus exposed to locational signals in these markets. Rather than relying solely on a capacity market with problematic availability and deliverability incentives or trying to shift directly to an energy only shortage pricing based market with reliability potentially compromised by regulatory risk, there is a potential middle ground of maintaining a basic framework of a locational capacity market but to gradually over time increase the importance of shortage pricing. This approach would limit regulatory risks because the capacity market would provide a stable revenue stream supporting an important portion of resource going forward costs, but would also tend to ensure deliverability and high availability because another important portion of the going forward costs would be recovered in day-ahead and real-time energy and reserve markets. It does not, however, address the incentive problems created by retail access programs; these must be confronted directly.

VI. CONCLUSIONS

Market-based resource adequacy mechanisms are inherently complicated by the fact that network reliability has elements of a public good and the probability of failure does not depend on the failure of individual market participants, which is allowed for in day-ahead and real-time security analysis, but results from too much failure by too many suppliers at the same time at locations that interact to produce an inability to reliably meet load. There is as a result inevitably a centralized element to analysis of reliability by decentralized consumers and resource suppliers. There is, therefore a tension between the centralized forward looking analysis of system reliability and the decentralized decisions of individual market participants. Both energy-only and capacity market resource adequacy mechanism have to use price systems, for energy or capacity, to provide appropriate incentives to decentralized market participants.

There is a transitional risk under either kind of approach that the price system (capacity or energy) as initially implemented will not provide appropriate incentives to decentralized suppliers and the intended level of reliability will not be supplied, or regulators will have to intervene and maintain the intended level of reliability with out of market interventions (RMR contracts). Beyond the problems arising from mistaken designs or transition issues, a resource adequacy system based on energy only pricing has the intrinsic feature that the spot market returns to capacity or conversely the cost to consumers of not being hedged with forward contracts will be concentrated in particular years. This concentrated impact gives rise to the potential for regulatory intervention to prevent the resulting wealth transfer from unhedged consumers to suppliers. The prospect of such intervention undermines the incentive of LSEs to contract forward for capacity and undermines the willingness of resource suppliers to build capacity not supported by long-term contracts (i.e., based on expected spot market revenues during shortage conditions).

While the energy only market outcome of very high prices during shortage conditions was allowed during 1998, in a market in which most consumers were hedged forward and the very high prices were borne by relatively small amounts of unhedged load or generators suffering atypical outages, this was not the case in California where some consumers (in San Diego) and large distribution companies (SCE and PG&E) had relatively low levels of hedging against price increases. Given the history of regulatory intervention through early 2006, it appears that there would be substantial regulatory risk associated with reliance on energy only markets to support resource adequacy, particularly in markets with retail access in which most consumers have very limited forward hedges against increases in the cost of power.

There may be mechanisms to reduce these regulatory risks, one perhaps being requirements for LSEs to enter into long-term forward hedging contracts covering a substantial portion (but not all) of their load. These alternatives will probably not be feasible in all regions, particularly in regions with retail access where there is no LSE with a long-term obligation to serve load on whom such a forward contracting obligation can be imposed.

The existing capacity market designs potentially provide a means of maintaining reliability despite this regulatory risk, by stabilizing the capacity component of payments for power over time. As discussed above, however, there is limited evidence for the proposition that capacity market will in practice operate in this manner.

In the long run, capacity market systems have several limitations relative to energy-only markets.

- *Deliverability Tests*: Because capacity providers are paid whether they are dispatched or not, capacity markets need rules to provide locational incentives.

- *Availability Standards*: Because capacity providers are paid whether they are available to be dispatched or not, standards are required for availability levels and timing.

- *Conservation Incentives*: Capacity market requirements are needed to keep extra capacity in operation because energy prices do not reflect the full cost of incremental consumption, so capacity markets need special rules to incent conservation.

- *Market Power:* The potential for the exercise of market power in long-run capacity markets is constrained by entry. Entry and exit, however, do not constrain daily or monthly capacity prices and retail competition leads to daily or monthly changes in capacity market requirements.

As discussed in Section III.I, experience to date with capacity market mechanisms in the U.S. suggests that they must be based on locational capacity markets rather than deliverability requirements and need to include performance incentives related to availability during reserve shortage conditions. Since only one locational capacity market is actually in existence and performance incentives based on real-time shortage conditions have yet to be implemented in any capacity market, the ability of these capacity market incentives to in general replicate the incentives of an energy only market is not yet established.

Because capacity markets designs suppress real-time prices during reserve shortage conditions they must attempt to provide efficient incentives for real-time demand response through capacity market incentives. These mechanisms generally entail paying consumers for not consuming. While most of the ISOs have made substantial efforts to provide incentives for demand response through their capacity market programs, their ability to provide efficient incentives is fundamentally compromised by the lack of real-time prices reflecting the actual reserve shortage costs, the ambiguities in measuring but-for consumption, and the potentially unbounded curtailment obligation incurred by demand response in a capacity market with reserve shortage based performance metrics.

While forward call contracts may be an attractive hedging mechanism in an energy-only market, requirements for forward contracting in the form of financial call contracts would be meaningless absent effective shortage pricing, such as in an energy-only market. Absent effective shortage pricing, requirements for call contracts supported by physical capacity are simply a more complicated and perhaps less transparent capacity market.

Moreover, none of these resource adequacy designs by themselves neutralize the effect of locational market power if it exists. Locational market power would often be effectively constrained through forward contracting supported by potential entry, but forward contracting in both energy only and capacity market designs is undermined by retail access programs, particularly those with provider of last resort or price to beat features. The adverse impacts of the lack of forward contracting by LSEs can be offset to a degree through forward capacity procurement by the ISO, but capacity procurement to meet load forecasts that are not tied to the demand for power at the resulting market prices carries risks of uneconomic forward capacity contracts and future stranded costs that will be assigned to consumers.

There are more problems and constraints in this resource adequacy problem than there are degrees of freedom in crafting a solution. Perhaps part of the long-term response by regulators must be to address some of the underlying problems such as the poor contracting incentives under retail access directly, through reform of retail access programs, rather than trying to solve these problems through the design of capacity or energy markets.

A second element of a strategy for addressing the unworkability of either energy only or capacity market may be to attempt to structure a middle ground in which a portion of the going forward costs of the marginal resource are recovered on a stable basis in a capacity market with some form of performance incentives related to availability during shortage conditions and another portion is recovered in margins during reserve or energy shortage conditions which while not set high enough to fully recover capacity costs (thus do not give rise to the extremely consumer cost volatility that produces regulatory risks), provides incentives derived from the energy market and ties a portion of capacity resource revenues and consumer costs to the years in which the system is stressed.

APPENDED TABLES

Table 41
Average Daily Natural Gas Delivered to Consumers by Region
(Excluding Vehicle Fuel) (MMcfd)

	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR
1997-1998									
CA	5,003	5,043	5,449	4,422	4,860	6,085	6,538	6,776	4,897
WA-OR	682	877	960	904	1,352	1,625	1,544	1,676	1,518
AZ-NV-NM	938	982	941	759	863	1,324	1,448	1,278	1,187
Total	6,622	6,903	7,349	6,085	7,074	9,033	9,529	9,730	7,602
1998-1999									
CA	4,759	5,315	5,415	4,900	5,153	6,200	6,912	7,142	5,613
WA-OR	927	1,119	1,127	982	1,383	1,536	1,714	1,617	1,388
AZ-NV-NM	1,108	1,153	968	934	1,046	1,490	1,467	1,372	1,184
Total	6,794	7,588	7,511	6,816	7,582	9,226	10,093	10,131	8,185
1999-2000									
CA	5,154	5,390	5,513	5,827	5,322	5,967	6,348	6,334	5,902
WA-OR	793	855	977	1,361	1,408	1,696	1,795	1,738	1,508
AZ-NV-NM	1,005	1,037	959	1,006	1,035	1,493	1,493	1,409	1,261
Total	6,952	7,281	7,448	8,195	7,765	9,156	9,637	9,481	8,671
2000-2001									
CA	6,031	6,817	6,184	6,250	6,658	6,896	7,860	7,713	6,494
WA-OR	1,058	1,103	1,154	1,256	1,449	1,638	1,762	2,130	1,708
AZ-NV-NM	1,270	1,461	1,289	1,230	1,494	1,776	1,741	2,010	1,635
Total	8,360	9,381	8,627	8,737	9,601	10,310	11,362	11,853	9,837

Note: EIA defines natural gas delivered to consumers as total consumption less lease and plant fuel and pipeline fuel.
Source: http://www.EIA.doc.gov/oil_gas/natural_gas/data_publications/;naturalgas_monthly/ngm.,html

Table 42
ISO-NE Capacity Prices

Month	Clearing Price ($/MW-Month)
Apr-98	$0.00
May-98	$0.00
Jun-98	$0.00
Jul-98	$0.00
Aug-98	$0.00
Sep-98	$0.00
Oct-98	$0.00
Nov-98	$0.00
Dec-98	$0.00
Jan-99	$0.00
Feb-99	$0.00
Mar-99	$246.00
Apr-99	$1,243.00
May-99	$523.00
Jun-99	$0.00
Jul-99	$0.00
Aug-99	$0.00
Sep-99	$0.00
Oct-99	$0.00
Nov-99	$0.00
Dec-99	$0.00
Jan-00	$1,250.00
Feb-00	$1,250.00
Mar-00	$1,250.00
Apr-00	$3,248.00
May-00	$2,500.00
Jun-00	$2,500.00
Jul-00	$2,500.00
Aug-00	$170.00
Sep-00	$170.00
Oct-00	$170.00
Nov-00	$170.00
Dec-00	$170.00
Jan-01	$170.00
Feb-01	$170.00
Mar-01	$170.00
Apr-01	$170.00
May-01	$170.00
Jun-01	$170.00
Jul-01	$170.00
Aug-01	$170.00
Sep-01	$4,870.00
Oct-01	$4,870.00
Nov-01	$4,870.00

Month	Clearing Price ($/MW-Month)
Dec-01	$4,870.00
Jan-02	$4,870.00
Feb-02	$4,870.00
Mar-02	$4,870.00
Apr-02	$4,870.00
May-02	$4,870.00
Jun-02	$4,870.00
Jul-02	$4,870.00
Aug-02	$4,870.00
Sep-02	$4,870.00
Oct-02	$4,870.00
Nov-02	$4,870.00
Dec-02	$4,870.00
Jan-03	$4,870.00
Feb-03	$4,870.00
Mar-03	$4,870.00
Apr-03	$400.00
May-03	$150.00
Jun-03	$200.00
Jul-03	$200.00
Aug-03	$230.00
Sep-03	$195.00
Oct-03	$120.00
Nov-03	$111.00
Dec-03	$87.00
Jan-04	$200.00
Feb-04	$10.00
Mar-04	$2.00
Apr-04	$30.00
May-04	$0.01
Jun-04	$6.00
Jul-04	$9.00
Aug-04	$10.00
Sep-04	$6.00
Oct-04	$0.02
Nov-04	$12.00
Dec-04	$25.00
Jan-05	$120.00
Feb-05	$700.00
Mar-05	$400.00
Apr-05	$175.00
May-05	$50.00
Jun-05	$100.00

Sources: NEPOOL Installed Capability Market Report (April 1998-March 2003), available at: http://www.iso-ne.com/settlement-resettlement/Pre-SMD_ICAP/Pre-SMD_Interim_ICAP/NEPOOL_Installed_Capability_Market_report.xls
Sources: NEPOOL Installed Capability Market Report (April 2003-Jloly 2005), available at: http://www.iso-ne.com/markets/othrmkts_data/inst_cap/icap/NewEngland_ICAP_Auction_Report.xls

Table 43
PJM 12-Month Rolling Average UCAP Capacity Payment
($/Year)

Month	12-Month Rolling Average Annual ICAP Payment: Daily Auction ($/Year)	12-Month Rolling Average Annual ICAP Payment: First Monthly Auction ($/Year)	12-Month Rolling Average Annual ICAP Payment: Last Monthly Auction ($/Year)
Dec-99	1,740.38	18,367.87	10,356.92
Jan-00	2,178.69	17,190.80	9,008.42
Feb-00	2,968.92	15,939.48	8,487.42
Mar-00	3,188.48	13,958.89	8,952.42
Apr-00	3,796.12	12,070.39	9,102.42
May-00	4,877.77	10,589.83	9,567.42
Jun-00	9,721.21	13,589.83	9,958.92
Jul-00	16,924.05	17,929.83	16,158.92
Aug-00	22,288.50	18,030.58	25,431.33
Sep-00	22,660.27	18,478.78	27,051.33
Oct-00	22,540.07	19,533.09	27,204.16
Nov-00	22,416.51	20,254.29	27,051.46
Dec-00	22,341.90	20,502.29	26,906.07
Jan-01	27,167.64	20,657.29	28,084.07
Feb-01	31,229.04	21,322.29	32,323.47
Mar-01	36,347.17	22,715.74	36,955.80
Apr-01	36,324.19	27,768.34	41,455.50
May-01	35,257.36	31,806.40	43,005.50
Jun-01	30,959.48	35,106.40	47,115.50
Jul-01	23,719.76	39,694.40	43,519.50
Aug-01	18,269.41	43,034.65	35,272.88
Sep-01	17,849.66	47,102.05	34,852.88
Oct-01	17,842.37	47,536.05	34,545.05
Nov-01	17,821.36	46,951.05	34,547.75
Dec-01	17,820.61	46,858.05	34,431.81
Jan-02	12,139.38	46,579.05	35,609.81
Feb-02	7,174.98	46,719.05	30,890.69
Mar-02	1,823.28	46,100.60	25,483.36
Apr-02	1,213.60	40,868.00	20,563.66
May-02	1,194.73	36,527.69	18,269.66
Jun-02	102.82	32,550.89	13,019.66
Jul-02	103.86	26,085.53	8,276.66
Aug-02	105.27	18,862.53	7,378.28
Sep-02	101.27	13,881.63	6,178.28
Oct-02	130.27	12,454.39	6,209.28
Nov-02	131.73	11,996.89	6,089.28
Dec-02	133.28	11,717.89	6,143.53

Table 43 (continued)
PJM 12-Month Rolling Average UCAP Capacity Payment
($/Year)

Month	12-Month Rolling Average Annual ICAP Payment: Daily Auction ($/Year)	12-Month Rolling Average Annual ICAP Payment: First Monthly Auction ($/Year)	12-Month Rolling Average Annual ICAP Payment: Last Monthly Auction ($/Year)
Jan-03	164.82	11,221.89	3,632.53
Feb-03	165.68	9,959.93	3,415.25
Mar-03	428.42	9,029.93	3,476.94
Apr-03	767.96	8,760.23	3,896.94
May-03	957.14	8,295.54	4,191.44
Jun-03	880.23	7,202.34	3,381.44
Jul-03	882.64	6,134.70	3,381.44
Aug-03	882.67	6,072.70	3,660.44
Sep-03	881.75	6,283.60	3,840.44
Oct-03	854.08	6,439.84	3,871.44
Nov-03	899.29	6,387.34	3,871.44
Dec-03	947.38	6,433.84	3,832.69
Jan-04	917.14	6,309.84	3,522.69
Feb-04	919.06	5,809.80	3,523.69
Mar-04	686.26	5,561.80	3,493.00
Apr-04	356.10	5,561.50	3,080.50
May-04	204.76	5,437.19	2,793.75
Jun-04	3,295.64	4,477.19	1,488.75
Jul-04	4,063.69	4,105.19	3,279.00
Aug-04	4,549.50	4,756.19	3,465.00
Sep-04	4,949.90	4,306.19	3,570.00
Oct-04	4,956.25	4,236.44	3,725.00
Nov-04	4,936.92	4,326.44	3,724.40
Dec-04	4,888.09	4,372.94	3,739.90
Jan-05	4,887.27	3,938.94	3,476.40
Feb-05	4,885.01	3,928.82	3,447.96
Mar-05	4,853.52	3,874.57	3,420.06
Apr-05	4,845.19	3,874.57	3,385.56
May-05	4,805.80	3,851.63	3,349.91
Jun-05	1,699.82	3,851.63	3,349.91

Bibliography

Addepalli, Raj, Harvey Arnott, Mark Reeder (New York PSC), Prepared Testimony Regarding a Proposal by the NYISO Concerning Electricity Capacity Pricing, March 6, 2003.

Bekman, Ken, FERC Docket ER 03-563-030.

Bidwell, Miles, "Reliability Options: A Market-Oriented Approach to Long-Term Adequacy," Electricity Journal, June 2005.

California Energy Commission, 2002-2012 Electricity outlook Report, February 2002.

Chandley, John, "ICAP Reform Proposals in New England and PJM," September 23, 2005.

Chandley, John, Scott Harvey and William Hogan, "Electricity Market Reform in California," November 22, 2000.

Charlton, John, "NYISO Demand Curve as Proposed for the NYCA Installed Capacity Market," March 14, 2002.

Cramton, Peter and Steven Stoft, "The Convergence of Market Designs for Adequate Generating Capacity," April 25, 2006.

Creti, Anna and Natalia Fabra, "Capacity Markets for Electricity," CSEM Working Paper, February 2004.

Dauphinais, Jim, "Reliability, Investment, Market Mitigation," PUCT Workshop in Resource Adequacy, June 17, 2005.

Dauphinais, Jim, "Why Energy-Only Markets?" PUCT Workshop on Resource Adequacy, June 17, 2005.

Desell, Art, Installed Capacity (ICAP), The Reliability Market.

De Vries, Lauren; H.P.A. Knops; R.A. Kakvoort, "Bilateral Reliability Contracts: An Innovative Approach to Maintaining Generation Adequacy in Liberalized Electricity Markets," May 25-27, 2004.

Goldman, Charles, Joseph Eto, and Galen Barbose, "California Customer Load Reductions during the Electricity Crisis: Did They Help to Keep the Lights on," May 2002.

Hamal, Cliff W. Answering Testimony, FERC Docket No. ER-563-030, November 4, 2004.

Harvey, Scott, M., William W. Hogan, and Susan L. Pope, "Comments on the California ISO MRTU LMP Market Design." California Independent System Operator, February 23, 2005.

Harvey, Scott M., William W. Hogan and Todd Schatzki, "A Hazard Rate Analysis of Mirant's Generating Plant Outages in California," January 2, 2004.

Harvey, Scott, M. and William W. Hogan, "Issues in the Analysis of Market Power in California, October 27, 2000.

Harvey, Scott M. and William W. Hogan, "California Electricity Prices and Forward Market Hedging, October 17, 2000

Hogan, William, "On an 'Energy-Only' Electricity Market Design for Resource Adequacy," September 23, 2005.

ISO-New England Inc., Initial Brief, Docket No. ER003-563-030, April 15, 2005.

ISO-New England Inc., Reply Brief, Docket No. ER003-563-030, April 27, 2005.

Internal NYISO DC Controllable Line Scheduling, Concept of Operations, May 4, 2004.

ISO New England, Manual for Installed Capacity.

ISO New England Load Response Program Manual.

ISO New England, Inc., Market Monitoring Department, "Interim Report on Electricity Supply Conditions in New England during the January 14-16, 2004 `Cold Snap'," May 10, 2004.

Jaffe, Adam B. and Frank P. Felder, "Should Electricity Markets Have a Capacity Requirement? If so, How Should it be Priced?", *Electricity Journal*, December 1996.

KeySpan Corporation, Motion to Intervene and Comments FERC Docket No. ER01-3001-014, January 23, 2006.

Karl, Mark, Prepared Direct Testimony, Docket No. ER03-563-030, August 31, 2004.

Klein, Abram (Edison Mission Marketing & Trading), "Ensuring Generation Adequacy in Competitive Electricity Markets," May 21, 2004.

Klein, Abram (Edison Mission Marketing & Trading), "Ensuring Generation Adequacy in a Competitive Electricity Market," June 18, 2002.

Klein, Abram (Edison Mission Marketing & Trading), "Do We Need ICAP? Ensuring Generation Adequacy in a Competitive Electricity Market," November 5, 2001.

Krapels, Edward, Paul Flemming and Stephan Conant, The Design and Effectiveness of Electricity Capacity Market Rules in the Northeast and California," *Electricity Journal,* October 2004

LaPlante, David, Prepared Direct Testimony, FERC Docket No. ER03-563-030, August 31, 2000.

LaPlante, David, Rebuttal Testimony, Docket ER03-563-030.

LaPlante, David, Supplemental Direct Testimony, November 4, 2004.

Meehan, Eugene, Chantale LaCasse, Philip Kalmus and Bernard Neenan, "Central Resource Adequacy Markets," Final Report, February 2003.

New York State Reliability Council, LLC, Installed Capacity Subcommittee, New York Control Area Installed Capacity Requirements for the Period May 2005 through April 2006, December 10, 2004.

New York State Reliability Council LLC, Policy No. 5-0, Procedure for Establishing New York Control Area Installed Capacity Requirements, August 11, 2003.

New York PSC, Resource Demand Curve, January 31, 2003.

NYISO Installed Capacity Manual.

PG&E National Energy Group, "Call Options for Energy: A Market Based Alternative to ICAP," October 2000.

PJM, PJM Reserve Requirements, Manual M-20, April 30, 2004.

PJM Manual, Capacity Obligations.

PJM Manual 21, Rules and Procedures for Determination of Generating Capability, April 30, 2004.

PJM Market Monitoring Unit, PJM Interconnection State of the Market Report 2001.

Oren, Shmuel, "Generation Adequacy via Call Options Obligations: Safe Passage to the Promised Land," UCEI Working Paper, September 2005.

Oren, Shmuel, "Capacity Payments and Supply Adequacy in Competitive Electricity Markets," May 21, 2000.

Rosellon, Juan, "Different Approaches to Supply Adequacy in Electricity Markets."

Ruff, Larry E., Affidavit, EX01-1.

Shanker, Roy, Affidavit, Docket No. EL05-1410-010, October 19, 2005.

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

State of New York Public Service Commission, Case 03-E-0641, Proceeding on Motion of the Commission Regarding Expected Implementation of Mandatory Hourly Pricing for Commodity Service, September 23, 2005 and April 24, 2006.

Stoft, Steven, Prepared Direct Testimony, FERC Docket No. ER03-563-030, August 31, 2004.

Steven Stoft, "Supplemental Direct Testimony," Docket ER03-563-030, November 4, 2004.

Vazquez, Carlos, Michel Rivier, Ignacio Perez-Arriaga, "A Market Approach to long-term security of supply," March 2001.

Woodley, John A. (Morgan Stanley), "Volatility, Capacity and Reliability."

Appendix A
Northeast RAM Process

PJM, NYISO and ISO-NE formed the Interregional Resource Adequacy Model Group (RAM group) to develop a coordinated approach to ICAP for the Northeast. The resulting RAM group model provided for centralized forward capacity purchases by the ISO, rather than by LSEs.[1]

The basic elements of the Central Resource Adequacy Model (CRAM) model were:

- Each ISO forecasts load and establishes an unforced capacity obligation for future operating years.
- Resources are committed to meet the unforced capacity obligation up to several years prior to the operating year.
- Each ISO coordinates its own separate centralized auction with coordinated timing.
- Auction participation (by sellers) is voluntary.
- Products that can be sold in the auctions include existing generation, planned generation, bilateral contracts for capacity resources, load management products and transmission upgrades.
- Bilateral contracts can be used by LSEs to self-provide their own generation.
- The centralized auction price of UCAP would be charged to all LSEs during the operating year.
- Resource providers would receive the market clearing UCAP price for their UCAP during the operating year.
- Periodic reconfiguration auctions to allow resource providers to cover changes in capacity positions, due to outages, unit cancellations, etc., but these later auctions would not change the price of UCAP for load, they would only price imbalances among suppliers.[2]

---

1 Joint Capacity Adequacy Group, Areas of Agreement and Areas Under Development, March 14, 2003 (hereafter JACG).

2 JACG; and Eugene Meehan, Chantale LaCasse, Philip Kalmus and Bernard Neenan, "Central Resource Adequacy Markets," Final Report, February 2003, (hereafter NERA), pp. 8-9.

NERA studied five issues relating to the CRAM model:

- Planning horizon (how far should the auction precede supply commitment?).
- Commitment period (length of supply contract awarded in the auction).
- Auction format (descending clock, reverse English, pay as bid).
- Proportion of ICAP acquired in each auction
- Deficiency charges and bid caps.[3]

Planning Horizon

The planning horizon issue addresses the question of by how long the UCAP auction should precede the date of the supply commitment. The longer the length of the planning horizon the more long-lead time resources can be offered in the UCAP auction. Conversely, if the planning horizon is too long, all resource projects other than existing generation will be so speculative that LSEs may be hesitant to commit to them and suppliers may be reluctant to commit to supply UCAP based on projects that are very early planning stages. Longer planning horizons would give rise to greater load forecast uncertainty that could increase UCAP requirements. It was also perceived that longer planning horizons would tend to exclude DSM programs from the ICAP auctions because consumers could not commit to demand-response programs that far in advance of the operating year.

The NERA Report proposed a minimum planning horizon of three years to allow entrants to compete effectively in the ICAP market. If the planning horizon were too short, capacity able to come on line in time to meet ICAP requirements would have to be so far along in the development/construction/financing process that the ICAP supply could not respond to UCAP prices. A short planning horizon would therefore tend to produce auction prices that were either zero or equal to the price cap, perpetuating current problems. NERA believed that it was desirable for the planning horizon to be long enough that development projects offered in the UCAP auction could go forward or not depending on the price in the UCAP auction.[4] NERA believed that a planning horizon satisfying this criterion would tend to stabilize UCAP prices around the long-run cost of generating capacity.

---

3 Joint Capacity Adequacy Group, Areas of Agreement and Areas Under Development, March 14, 2003. NERA, p. 10.

4 NERA, p. 13, 15—22.

One potential problem with longer planning horizons is that since projects offered in the auction would necessarily be at an earlier stage in the development process than would be the case with a shorter horizon, there would be a greater potential for the projects to turn out to be uneconomic or infeasible for reasons not known at the time of the capacity auction. This possibility would give rise to non-performance risk for the ISO running the auction as developers could submit speculative offers, anticipating that they would cancel the project if fuel costs or market prices moved in an unfavorable direction.

This kind of non-performance risk could be addressed by requiring that winning bidders post some kind of security, but such a security requirement could deter participation in the auction if the planning horizon were too long and the auction occurred too early in the development process for most projects. Indeed, these kinds of credit issues were a market participant concern with the CRAM proposal. NERA proposed to limit non-performance risk by restricting participation in the auction to projects in an advanced stage of development, siting and permitting,[5] but this view did not accord with the perspective of other capacity market supporters.

In considering these risks, it needs to be kept in mind that under the current capacity market systems the capacity requirements are only enforced a month in advance, so the possibility of projects being cancelled or delayed by environmental or other factors under the CRAM system between the time of the initial auction and the operating month does not create any new reliability risk relative to current capacity markets. Moreover, although the NERA study did not propose this, defaulting capacity could be replaced in the reconfiguration auction.

NERA concluded that the three-year planning horizon appeared to be inconsistent with the participation of demand response resources and proposed carving out a portion of the capacity market to be served in a shorter term auction for demand response resources.[6] This proposed approach was not popular with market participants because of the likelihood that this would in effect create separate capacity markets and increase price volatility in the portion of the capacity market served by demand response. An alternative approach would be to allow resources that did not sell capacity in the three year forward auction to participate in the reconfiguration auctions and to allow offers in the three year auction that were not backed by explicit generating resources or demand response resources but reflected an expected ability to develop such resources. The non-performance credit cost issue associated with such offers could perhaps be addressed be requiring stringent credit requirements only for capacity offers not backed by projects or for projects in a very early stage of development.

---

5 NERA, p. 21.

6 NERA, pp. 31-32, 141-142.

Commitment Period

The commitment period issue concerns the duration of the capacity contracts awarded in the auction. Longer contract durations reduce the risk for the seller and lower the price of capacity but entail a longer financial commitment by the ISO.[7]

NERA recommended contracts with a duration of three years due to their assessment of the limits of centralized auction and customers. NERA proposed to implement this three-year commitment period by running annual auctions covering one-third of the capacity requirement purchasing three-year capacity requirements. Thus, in 2005 the ISO would purchase one-third of the capacity requirement for 2009, 2010 and 2011; in 2006 the ISO would purchase one-third of the requirement for 2010, 2011 and 2012; in 2007 the ISO would purchase one-third of the ICAP requirement for 2011, 2012 and 2013, etc.[8]

A three-year contract duration provides a desirable improvement over the current monthly duration or capability period duration of capacity contracts in the centralized auctions. Nevertheless, a three-year contract may not provide a sound basis to support the construction of new capacity and thus may have little practical impact on resource development.

An important feature of the CRAM proposal is that it can be bypassed by LSEs desiring to enter into longer-term customized contracts with resource suppliers. Such LSEs could sign 20-year contracts covering capacity as well as energy and offer one-third of the capacity they procure into each auction, hedging themselves against the capacity charges. Thus, the three-year contract term in the centralized auction does not preclude longer-term bilateral contracts. As noted above, however, LSEs may not have an incentive to enter into such long-term contracts in states with shorter-term retail access programs.

The CRAM proposal in effect served mainly to ensure that someone (the ISO on behalf of future loads) contracts for capacity to cover the portion of load that may migrate from LSE to LSE and thus would not likely be hedged under long-term bilateral contracts. Moreover, the capacity costs of serving this load would be known well in advance, enabling these costs to be recovered in retail contracts.

Single Auction or Tranches

One alternative would be to acquire the entire capacity requirement for a given year in a single auction. This was perceived to have the advantage of yielding a single price.

---

7 NERA, pp. 22-25, 28.

8 NERA, pp. 13, 22-25.

Another alternative would be to stagger the acquisition of capacity, for example acquiring one-third of the requirement one year, one-third the next and one-third the next. NERA recommended a staggered approach as better minimizing market power problems and accommodating the auction of multi-year capacity streams.[9]

Auction Format

NERA proposed a Descending Clock Auction. NERA viewed such an auction format as good for inter-ISO procurement and for taking account of imports.[10]

Deficiency Charge and Bid Cap

The issues were whether there ought to be a bid cap in the auction and how the deficiency charge, for suppliers that failed to perform, should be set.

NERA recommended a liquidated damages approach to setting deficiency payments, perhaps higher in the winter than the summer. NERA recommended that deficiency charges be assessed and come due at the time a capacity provider's failure to perform was known, suggesting that missing a construction milestone and abandoning construction shortly after winning the auction would trigger the charge. The same penalty would apply if a unit's capacity rating fell below the level sold in the auction.[11] This mechanism was tied in NERA's proposal to qualification criteria limiting participation to units in an acceptable stage of development and tied to a specific physically verifiable plant at a specific location (i.e., no virtual capacity supply bids).[12] The triggers proposed by NERA do not appear consistent with permitting capacity sellers to cover shortfalls by purchasing power in the reconfiguration auctions as envisioned by the original CRAM proposal.

Some market participants and regulators envisioned the CRAM proposal as permitting virtual supply offers.

Other Issues

The proposed three-year auction structure raises a variety of market power mitigation issues, some of which are discussed in the NERA report.[13]

---

9 NERA, pp. 37-58.

10 NERA, p. 13.

11 NERA, p. 83.

12 NERA, p. 84.

13 NERA, Section 7.

Another important limitation of the NERA/CRAM proposal, particularly from the standpoint of NYISO market participants, was that it did not include an ICAP demand curve. This appears to have been a major stumbling block with Northeast market participants (and regulators). The capacity demand curve has been to be relatively popular with NYISO market participants, many of whom believed the ICAP demand curve in New York to be working much better than any of the other ICAP mechanisms and were reluctant to abandon it for an untried design.

NERA took the view that an capacity demand curve was compatible with the proposed CRAM auction structure.[14] NERA criticized the concept of an capacity demand curve on a number of grounds, some of which are consistent with the comments above.[15] As discussed above, an capacity demand curve appears to have important potential limitations, but most have yet to manifest themselves. The CRAM approach would avoid the outcome of meaningless capacity auction prices if generation projects would go forward or not based on the price of capacity over a three-year term, as low prices would back out investment until prices rose and high prices would draw in investment. If a three-year contract term is not meaningful for project financing, however, then the CRAM proposal will not end the cycle of capacity auction prices that are either zero or equal to the deficiency payment.

The CRAM proposal has a few other potential implementation issues or limitations:

- With a three-year planning horizon, how would load forecast errors that result in an ICAP shortfall be addressed?
- What happens if the ISO qualification process excludes resources that have already entered into bilateral capacity contracts?
- How would simultaneous but separate capacity auctions be coordinated?

A key feature of the CRAM proposal was that it is intended to address a very limited set of issues. It does not address the issues relating to deliverability requirements, outage incentives, or unit availability that were discussed above.

---

14 NERA, pp. 119-126

15 NERA, p. 130-136.