Review of the Reserves and Operable Capability Markets: New England's Experience in the First Four Months

Peter Cramton[1]

Professor of Economics, University of Maryland
President, Market Design Inc.

November 17, 1999

I have been asked by ISO New England to provide a review of the performance of the operating reserves and the operable capability markets. Here I provide a review of the three reserve markets—ten-minute spinning reserve (TMSR), ten-minute non-spinning reserve (TMNSR), and thirty-minute operating reserve (TMOR)—and the operable capability market (OpCap). The review covers the first four months of operation from May 1 to August 31, 1999. My review is based on my knowledge of the market rules and their implementation by the ISO, and the market data during this period, including bidding, operating, and settlement information. Since that data are confidential, I have presented only aggregate information in the tables and figures that follow.[2] Although this review will cover only the reserves and OpCap markets, I have reviewed the data from the energy and AGC markets as well. Since all of the NEPOOL markets are interrelated, one cannot hope to understand one market without having an understanding of the others.

In this review, I hope to accomplish four tasks:

1. Identify the potential market flaws with these markets.

2. Look at the performance of the markets to see if the potential problems have materialized.

3. Evaluate the ISO's short-term remedies for these market flaws.

4. Propose alternative medium-term solutions to the identified problems.

Each of these tasks is substantial. In order to provide this report within a tight schedule, I have limited the review to the major issues. In particular, my analysis of the data is limited, but I believe sufficient to draw first-order conclusions. Also, I ignore issues stemming from operational problems at the ISO. Not surprising, the ISO has experienced some operational problems, given the complexity of the markets and the tight timelines for implementation. Although these problems have impacted the reserve and OpCap markets, I do not dwell on them here, since it is my understanding that the ISO is already working hard on solving these problems. Finally, I do not look at the ICap market. The ICap market shares many of the flaws with the OpCap market; however, it has proven to be less of a problem, since bids are submitted on a monthly, rather than a daily basis. A more comprehensive study of the markets will have to wait for a later date.

---

[1] The views expressed are my own. Please send comments to cramton@umd.edu. Related work can be found at www.cramton.umd.edu and www.market-design.com. I am grateful to Jeffrey Lien for outstanding research assistance, and to Hung-Po Chao, Alvin Klevorick, Robert Wilson, ISO New England, and numerous NEPOOL participants for helpful comments.

[2] When particular bidders or units are discussed it is by a bidder or unit number that I have created. No cross-reference to an actual bidder or unit ID or name is provided.

I conclude that the reserves and OpCap markets are seriously flawed. Looking at the data, one might argue that most of the time these markets are working well; that is, prices are "reasonable" most of the time. But this is simply a reflection that most of the time these markets are irrelevant—most of the time the resources these markets price are not scarce. In the absence of scarcity, the markets are doing a reasonable job of pricing and assignment. But this is similar to an air conditioner working well in the winter. It works well, simply because it is not asked to do anything. Like an air conditioner, the reserves and OpCap markets should be evaluated when they are stressed—when the markets are actually resolving a problem of scarcity. On this score, the markets have performed poorly (although much better than the early months in some other markets, such as California).

The poor performance of these markets is not a surprise. It is the result of basic market flaws. These flaws have been recognized by the ISO and NEPOOL participants, since well before the markets began operation. Unfortunately, there was insufficient time to fix these markets before operations begin on May 1, 1999. It was felt that the ISO would be able to introduce short-term fixes as needed in the early months of operation. This has worked reasonably well. The flaws have not been fatal. The ISO has been able to operate a workably-efficient and reliable wholesale electricity market. Indeed, it is remarkable how well the markets have worked, despite their flaws. One measure of this is that the cost of reserves is still a small percentage of total energy costs.[3]

Ideally, what is needed is a long-term fix to the basic design flaws in the reserves and OpCap markets. NEPOOL's Congestion Management and Multi-settlement Committee is in the process of developing a long-term solution. Realistically, it will not be possible to implement the long-term solution until some time in the year 2000, at best. In the mean time, it is worthwhile to identify a medium-term fix that the ISO can implement within a matter of months. I present such a medium-term solution below.

Recommendation: Eliminate the OpCap market.

Given the nature of the market flaws and the difficulties of implementing more than a modest variation of the status quo, I believe the best approach is to eliminate the OpCap market. It is hard to argue that the OpCap market serves any useful purpose as presently designed. Since its operation has real costs with no apparent gain, it is best to eliminate the market. Until the market is discontinued, the ISO's limits on prices should be maintained.

Recommendation: Adopt the smart buyer model for the reserve markets.

I recommend that the reserve markets continue in the medium-term, but with the ISO adopting a smart buyer model for reserves.[4] As a smart buyer, the ISO:

1. Never pays for additional reserves more than the economic value of the additional reserves.

2. Reduces its demand for reserves as reserve prices increase.

3. Shifts purchases toward higher quality reserves when they are priced less than lower quality reserves.

---

[3] In July, the worst of the four-month period, reserves were 3.6% of the total energy bill. In contrast, in California's first summer of operation, prices in the reserve markets were often chaotic and often substantially above energy prices. California's ancillary service costs in July 1998 averaged $21/MWh, when energy prices averaged $32/MWh. California's ancillary service cost before market-based rates averaged just over $1/MWh.

[4] The smart buyer model is related to, but different from, the Rational Buyer Model implemented by the California ISO.

In conducting the markets for reserves, the ISO is effectively purchasing reliability on behalf of load. The ISO has a responsibility to purchase wisely, given the aggregate preferences of load. On this basis, the ISO as smart buyer develops a demand curve for reserves that reflects the marginal value that additional reserves have for the system.

The essential elements of the smart buyer model can be adopted in the medium-term. The ISO recently has implemented point 3 above. The reserves now are treated as a cascade with the ISO filling its need for lower quality reserves with higher quality reserves if the higher quality can be purchased at a lower price. The difficult step is establishing a demand curve for reserves, which reflects the marginal value of additional reserves. This will accomplish points 1 and 2 above. It will require some discussion among participants and regulators, but once new operating procedures are established the ISO can begin purchasing reserves according to its demand curve, rather than the current vertical "demand" curve, which does not reflect the marginal value of additional reserves. Details of how this might work are discussed in Section 5.

Recommendation: Restructure the reserve markets.

The basic flaws in the reserve markets stem from two features of the current markets:

1. The vertical "demand" curve, derived from a rigid reserve requirement, implies that prices are arbitrarily high in times of scarcity.

2. Losing bidders face the same obligations as winning bidders.

The first flaw is eliminated by constructing the true demand curve for reserves, based on the marginal value of additional reserves. The second flaw is solved by paying everyone that provides reserves in real time the market clearing reserve price. The real time supply curve is calculated by the ISO based on the state of the system. The supply corresponds to the quantity of reserves that is available in the system. The clearing price is then found at the intersection of this vertical supply curve with the demand curve for reserves. All units providing reserves are paid this clearing price.

Recommendation: Create a multi-settlement energy and reserve markets as soon as possible.

In the long term, the reserve markets should be restructured. NEPOOL and ISO New England should continue to work aggressively at designing and implementing multi-settlement energy and reserve markets. The medium-term solution that I propose is an important step in moving toward for these markets is not a permanent fix. It does not make sense to purchase the entire reserve requirement in real time. The real-time market should simply be a balancing market to acquire and price deviations from the day-ahead scheduled reserve plan.

The report is organized as follows. Section 1 identifies the potential market flaws. Section 2 demonstrates that the flaws were observed in each of these markets. Section 3 describes the ISO's short-term fix of the flaws, and evaluates whether the ISO's remedy has been successful. Section 4 presents an alternative remedy that could be implemented in the medium-term. Section 5 discusses how the ISO could implement a downward-sloping demand curve for reserves. Section 6 describes a long term solution to the reserve markets, and demonstrates that the medium-term solution is an important step toward the long-term solution.

1   Potential market flaws

The energy and reserve markets represent the core of the system; there is no separate market for transmission, although one is under development. These markets cannot be viewed in isolation, as each

interacts with the others. In what follows, I examine the markets as a system, recognizing any interdependencies. However, the discussion focuses on the three reserve markets and the OpCap market.

The Table 0 below presents the salient features of the seven markets.[5] The three reserve markets and OpCap are structured in a similar way, and hence all four suffer from the same basic problems.

Table 0. Salient Features of the NEPOOL Markets

Market	Product	Residual or full requirements	Bid Submission	Settlement (all markets are settled after the fact)	Cost Burden	Losers provide the service?
Energy	Electrical energy in MWh	Residual	Hourly bids submitted day-ahead	Single hourly clearing price Out-of-merit-order suppliers paid based on their bids	Load	No
Automatic generation control (AGC)	Automated load following in regs	Full requirements	Hourly bids submitted day-ahead	Single hourly clearing price plus payment for AGC actually provided	Shared Proportionally by load	Yes
Ten-minute spinning reserve (TMSR)	Reserves that are synchronized to the system and capable of responding within ten minutes in MW	Full requirements	Hourly bids submitted day-ahead	Single hourly clearing price includes lost opportunity cost component	Shared proportionally by load	Yes
Ten-minute non-spinning reserve (TMNSR)	Reserves that are capable of responding within ten minutes in MW	Full requirements	Hourly bids submitted day-ahead	Single hourly clearing price	Shared proportionally by load	Yes
Thirty-minute operating reserve (TMOR)	Reserves that are capable of responding within thirty minutes in MW	Full requirements	Hourly bids submitted day-ahead	Single hourly clearing price	Shared proportionally by load	Yes
Operable capability (OpCap)	Operable capacity of each participant in MW	Residual	Hourly bids submitted day-ahead	Single hourly clearing price based on bids of participants with excess operable capacity	Participants who are deficient pay those with excess	Yes
Installed capability	Installed capacity of each participant in MW	Residual	Monthly bids submitted day before month starts	Single monthly clearing price based on bids of participants with excess installed capacity	Participants who are deficient pay those with excess	Yes

1.1   *Losing bidders face the same obligations as winning bidders*

The critical distinction between the energy market and the reserve markets is in the last column of the table above: Losers provide the service? In the energy market, the answer is no. Only the winning bidders provide the service (energy): (1) bidders submit supply schedules, (2) a market clearing price is determined from the intersection of the aggregate supply schedule and the realized demand, (3) bids at or below the clearing price are accepted, and (4) the winning bidders receive the clearing price for product (energy) delivered. However, in the reserve and OpCap markets, *there is no difference in the costs or risks incurred by those participants who receive payment in the market and those who do not.* Every participant is providing the same service, but only those designated are paid. As a result the only rational bids in the market are a bid of zero (to insure selection in the hope there is any positive price) or a bid that is an attempt to set the clearing price. The winning bidders are receiving payment for product delivered, but the losing bidders are delivering the product as well without receiving any payment.

This paradox that losing bidders incur the same costs as winning bidders stems from two features of the current markets. The first is the obligation of generators to respond to dispatch instructions. The second is the single-settlement system. Since the market is settled after the fact, the generators do not

---

[5] Further details are provided in the appendix. However, for a detailed description of the markets one should look at the market rules, which are available at www.iso-ne.com.

know until real time whether they are providing energy, and they do not know whether they were designated for reserves until after real time.

This problem was recognized in the March 5, 1999, Multi-settlement Proposal, which was approved by NEPOOL and filed with FERC in March 1999. Central to the current markets is the dispatch obligation of operable units. Both committed units and offline units are obligated to respond to ISO instructions.[6] Hence, the amount of TMSR in the system is the unloaded capacity of online units that can be ramped in 10 minutes, the amount of TMSR+TMNSR is the unloaded capacity of online and offline units that can be ramped in 10 minutes, the amount of TMSR+TMNSR+TMOR is the unloaded capacity of online and offline units that can be ramped in 30 minutes. In real time, the ISO draws on the energy stack in merit order to balance the system. The ISO also designates in real time certain unloaded capacity as TMSR, TMNSR, or TMOR, based on the reserve offers. These designated resources are paid the corresponding reserve clearing price.

The problem is that all rampable units are providing dispatch flexibility, yet only the designated one's are paid for it. This distorts the bidding. Regardless of what it costs to provide reserves, a bidder is better off being paid than not; hence, if it does not think it will determine the price, it should bid 0 to maximize the chance that it will be paid for reserves. Since others will do likewise, the equilibrium price is biased toward 0, assuming no market power.

The markets do not give the participants a meaningful way to express the costs they incur in providing dispatch flexibility. In a real market, the winning bidders would be paid for the product delivered (dispatch flexibility), but losing bidders would not be forced to deliver the product as well. A solution to this problem is presented in Section 5.

1.2   In times of scarcity, prices in these markets are arbitrarily high

Four of the markets, TMSR, TMNSR, TMOR, and OpCap, are particularly vulnerable during periods in which all or nearly all available resources must be selected to meet the reserve requirements. In situations where the reserve requirement cannot be met (Operating Procedure 4 conditions), prices may be arbitrarily high with no basis in cost and no economic constraint on bid behavior. In these situations, there are insufficient bids to satisfy the requirement. The ISO must accept all bids.

The auction becomes equivalent to the game of "ask and it shall be given." In this game, the auctioneer asks each participant to write a number on a piece of paper (a "bid") and agrees to pay each person selected the number of dollars bid by the person with the highest bid selected. If there are ten bidders and the auctioneer announces that seven of the ten will be selected in each round, there is a pressure to drive the prices to zero, even if there are real costs associated with participation. However, if the auctioneer announces in advance that all ten will be selected, the only limit on the bids is the auctioneer's bankruptcy. The forecast by the ISO of OP4 Conditions is equivalent to announcing that all bids in the reserve and OpCap markets will be selected.

Faced with these incentives, certain participants have taken advantage of this vulnerability by submitting low bids in the OpCap market for most of the capacity of their units and high bids for one megawatt blocks near their units' High Operating Limit. With this strategy, the bidder receives the clearing price on as large a quantity as possible when there is excess supply, and then in times of scarcity sets a very high clearing price. If others follow this sensible strategy, the OpCap price is either near zero

---

[6] For example, the ISO has the power to curtail external transactions based on reliability considerations. This command-and-control approach is viewed as necessary in the medium-term to maintain reliability. In the long run it should give way to a market-based approach, whereby curtailment is done on an economic basis and those that are curtailed are appropriately compensated.

or arbitrarily high, depending on whether there is excess supply. One cannot argue that this game of "ask and it shall be given" is an efficient means to compensate participants for offering capacity in times of shortage. Surely, well-designed energy and reserve markets are a better instrument to reward those that offer scarce capacity.

In a competitive market, when there is a shortage of supply, prices are determined from the aggregate demand curve. That is, in times of shortage, buyers respond to the higher prices by demanding less, which limits any price increases. Unfortunately, in the current reserves and OpCap markets such a market response is not effective—demand for reserves and OpCap is completely inelastic. Hence, in times of shortage, there is no market constraint on what suppliers can bid. The point is that the bids are not reflective of any costs. The price is set at the whim of the bidder willing to submit the highest number. This is a clear case of market failure.

The basic problem is the absence of a demand curve for reserves, which reflects the marginal value of additional reserves. A solution is suggested in Section 5.

1.3   OpCap provides a service of no value

The OpCap market, which shares the problems above of the reserve markets, has an additional problem. OpCap provides a service of no value. OpCap is a holdover from an electricity market with regulated prices. While it had a purpose under regulated pricing, it has no role in a competitive electricity market. All the electricity market needs are well-designed energy and reserve markets. The OpCap market does not contribute in any way.

OpCap is best thought of as an option. The ISO is buying an option to call the resource to provide energy or reserves in real time. In valuing this option, the critical number is the "strike price," which in this case is the (minimum) price the ISO must pay the resource when it calls it to provide energy or reserves. The difficulty is that the strike price has no impact on whether the resource gets designated. It is designated solely on the basis of its OpCap bid. Hence, the resource can make the option worthless by bidding extremely high energy and reserve prices. An analogous situation would be my selling you an option to buy 100 shares of Microsoft stock tomorrow. How much would you be willing to pay for this option if you knew I would set the strike price after you offer me a price for the option? Of course, I have an interest in setting an extremely high strike price, giving you the option to buy the 100 shares, but only at an extremely high price. Such an option has a value of 0. Similarly with the OpCap market, OpCap only has value if what is being offered is capacity that can be called at reasonable energy and reserve prices. Since the generators are making no such commitment, OpCap has zero value.

One solution is to tie the designation of OpCap to the energy and/or reserve bids. Such an approach demonstrates the redundancy of the OpCap market with the reserves and energy markets. What the ISO needs to run the system is energy and dispatch flexibility. The ISO is better off buying these products directly in well-designed energy and reserve markets, rather than buying them indirectly in an OpCap market. At best the OpCap market is redundant. But more likely it is destructive to efficiency, since it is likely to distort bidding in the energy and reserve markets. Worse yet, it can stand as an entry barrier, since only certain capacity is able to supply OpCap. For example, a new entrant is only eligible to supply OpCap after certain conditions have been met. Also imports are not able to provide OpCap. Finally, OpCap discourages entry is by rewarding nearly-mothballed capacity. This capacity may be able to limp along on OpCap and ICap payments, and not provide any useful service (it may submit extremely high bids in energy and reserve markets with long minimum run times and little ramping capability). Its presence discourages the entry of more useful capacity.

2   Performance of the markets

I have conducted a preliminary analysis of all the bidding, pricing, and settlement data from May through August 1999. The data are most easily described in summary figures and tables. These are contained in the companion file Cramton-on-Reserves-and-OpCap-May-Aug-1999-Exhibits.pdf. The exhibits are best viewed on a large monitor or printed out on a high-resolution color printer (high-resolution black-and-white is a reasonable alternative). In any case, the reader should have the exhibits available when reading this section.

2.1   Understanding the figures and tables

Figures 1–11 show the evolution of Clearing Prices and System Load from the opening of all markets on May 1 through the end of August. Each graph shows a two-week period. Figures 1 and 2 both plot the Energy Clearing Price and System Load over time—the difference is that the price scale of Figure 1 extends to $1000/MWh, whereas the price scale of Figure 2 extends only to $50. Figure 1 illustrates the timing and magnitude of the major peaks in the energy market. Figure 2 shows the day-to-day peaks. All reserve prices are plotted similarly on multiple scales in consecutive figures. The graphs with a restricted scale cut off the highest peaks, but show the more common hourly changes. The graphs with the unrestricted scale show the dramatic, but infrequent major peaks.

Figures 12 and 13 show the price duration curves for each of the four months under consideration. Each row represents a different price (Energy, TMSR, TMNSR, and TMOR), and each column represents a different month (May, June, July, and August). For example the graph in the upper left-hand corner of Figure 13 is the duration curve of the energy price in May. It shows that in May the energy price exceeded $20 nearly 100% of the time, but almost never exceeded $40. Figures 12 and 13 only differ in scaling. Notice that the fraction of hours axis in Figure 12 extends only to 0.3. This allows a detailed look at what happened when prices were peaking; Figure 13 presents the entire duration curve.

Figures 14–16 show how the prices of Energy, TMSR, TMNSR, and TMOR were correlated with load. Again each graph represents one month and one market. For example the graph in the lower right hand corner of Figure 15 shows how the TMOR price was correlated with System Load in August. The graph shows that only when load reached its highest levels did the TMOR price differ much from 0. Figure 14 plots all the prices that occurred from May through August; Figure 15 only plots prices at or below $200; and Figure 16 plots prices at or below $50. The different scales enable the reader to focus on different types of price changes, from the dramatic (Figure 14) to the common-place (Figure 16).

Figures 17–20 are graphs of average prices for each hour of the day for each month. Figure 21 shows these same averages for the entire period of May through August. The number of the hour on the x-axis is the ending time of the hour in question. For example, in Figure 19 the TMSR price graph (the graph labeled with triangles) has a spike above hour 23. This demonstrates that TMSR prices tended to be high in July for the hour ending at 11 pm.

Tables 22–23 provide summary statistics of market conditions. The tables provide overall summary statistics, statistics broken down by month, and statistics broken down by month and peak versus off peak. Peak hours were defined with the help of Figure 21. For all peak hours (hour 09 through hour 22) average load from May 1 to August 31 exceeded 15,000 MWh.

2.2   Load

Throughout the month of May load remained consistently low. The maximum value of 15,799 MWh was quite small compared to the peaks that were to come in June and July. Early in June New England experienced two significant heat waves. Hourly load rose to 18,375 MWh on June 1, 19,013 MWh on June 2, 20,866 MWh on June 7, and 20,794 on June 8. The June 7 load levels were particularly

unexpected. By 8 AM on June 7 actual demand was already nearly 900 MW above forecast demand. The load peaks of June 28 and 29 were predicted. Hourly load reached an all-time record peak of 21,840 MWh on the 28th and 21,311 MWh on the 29th. The forecast for the 29th had been more ominous than the actual realization of demand. The highest load of the summer, 22,426 MWh, was recorded on July 6 between 3 and 4 PM. Nine other days in July saw hourly load levels exceed 20,000 MWh with 4 of these days including load levels above 21,000 MWh. Load levels saw a slight decline in the month of August. The 21,000 MWh mark was not reached in any hour of August, although 20,000 MWh was surpassed numerous times. The average hourly load was slightly higher in August than it had been in June, but the peaks in August were not as intense.

2.3 Market performance

Prices in the reserve markets behaved predictably. Prices were close to or at zero when system conditions resulted in excess supply of reserves. This was the case for most of the month of May (when load was low), and for most of the month of August (when system capacity was high). Prices were arbitrarily high when system conditions made reserve supplies scarce. June and July were both characterized by low reserve prices for the majority of hours, but wildly volatile price spikes on the days with the highest load levels.

2.3.1 May

As expected the reserve markets opened quietly. The relatively low load levels resulted in excess availability of all reserve types throughout the month of May. The TMNSR price and the TMOR price both stayed below $1 for the entire month. The TMSR price hit peaks of $10.65 May 19, $12.63 May 25, and $46.44 May 31, but remained below $1 for almost the entire rest of the month.

2.3.2 June

The first true test of the reserve markets came on June 1 and June 2, the first high load days of the season. The high load reduced the reserves available for the TMNSR and TMOR markets. The TMOR price broke the one-dollar barrier for the first time on hour 12 of June 1. Just two hours later the TMOR price hit $129. At the same time, the TMNSR price was $125, and the TMSR price was $0.80. This complete inversion of prices, with higher prices given to inferior reserves, persisted for much of June 1 and June 2. The TMSR price rose slightly on June 2, but did not rise excessively until June 7 and 8 when it exceeded $400 for several hours, and eventually surpassed $800. June 7 and 8 were also the first days that the energy price hit major peaks. Unlike the TMNSR and TMOR prices, the TMSR price includes an opportunity cost payment, which is based on the energy price. This opportunity cost component causes the TMSR price to increase significantly whenever the price of energy peaks. The TMNSR and TMOR prices also rose on June 7 and 8, but not nearly as much as the TMSR price. Figure 12 shows that although prices above $100 dollars were not uncommon in the TMNSR market during the month of June, the price never exceeded $150.

Table 26 shows that on June 1, June 2, June 7, and June 8 the TMNSR price and the TMOR price regularly exceeded the energy price. Paying more for reserves than energy does not lead to rational cost minimizing procurement of electrical services. The ISO took steps to correct this market flaw on June 27 when it implemented a cap on the reserve prices (see section 3). The cap resulted in revised TMSR prices for June 27 and 28. The original operations data reported that the TMSR price reached $999/MWh on hour 12 of June 28 and $814/MWh on hour 13. When the cap was implemented the TMSR prices used for settlement were revised to $308/MWh and $489/MWh respectively.

2.3.3 July

July 6 experienced record prices in the TMSR market in addition to record load levels. The price of TMSR climbed to nearly $1000/MWh and remained above $700 for 6 hours. Originally energy prices also were set close to $1000/MWh for these hours, but these prices have since been revised to $500/MWh and lower. The TMNSR and TMOR prices continued their pattern of escalating with high load, but not excessively so. TMNSR and TMOR hit peaks of $116/MWh and $90/MWh respectively on July 6. The TMSR price should be higher than the inferior reserve prices, but a gap of over $800/MWh cannot be justified.

Figures 4, 6, 8, and 13 show that although there were many spikes in the reserve prices throughout June and July, the prices between the spikes stayed very low. Table 13 shows that in all three reserve markets the price stayed below $5 for at least 80% of the hours each month in all 3 markets. This price behavior is to be expected. The capacity of the system reaches its maximum during the summer months, so when load is not peaking there is plenty of capacity available for reserves.

2.3.4 August

Figures 10, 15, and 16 demonstrate that all markets settled down significantly in August. No price exceeded $100/MWh in the entire month of August, and prices in the TMNSR and TMOR markets returned to 0 almost as frequently as they had in May. Part of this is due to an overall decline in load, but even when load was high, prices did not peak like they did in June and July. Figure 26 shows that reserve prices, particularly the TMSR price, often reached the energy price cap during the second half of August, however this is somewhat misleading. Of the 22 hours when the energy price equaled the reserve price, 19 were in hours that have been labeled off-peak (hour 23 through hour 08). In June and July the average energy price in hours when the TMSR price equaled the energy price was $192/MWh, whereas during August the average energy price in this type of hour was only $27/MWh. Chances are that although the cap had an effect on the TMSR price in many hours of August, the effects were not very big. TMSR prices often rise during late night hours as units shut down and availability falls, but the price increases tend to be small (see Figure 20). Figure 16 demonstrates that during the month of August the TMSR price was often high when load was low. This seemingly puzzling result is because of these late night price increases. Of the 46 hours in August that the TMSR price exceeded $10/MWh, 22 were between 11 pm and 1 am. These late night price spikes also exist in the other months (particularly for the hour ending at 11 pm), but when compared to the peak hour price spikes they appear small.

The return to low prices in August is best explained by the increase in total capacity available, not by the cap on prices. Figure 29 shows the sum of High Operating Limit and Low Operating Limit bid in each hour from May through August. The upward trend of both graphs indicates the increase in available capacity as the summer progressed. Often in June and July, System Load exceeded the sum of High Operating Limit and the market was forced to search for imports. This never happened in August.

2.4 Inefficient designations

Reserve designations are often assigned in such a way that the total cost of supplying reserves is not minimized. Inefficiencies arise because the markets are cleared sequentially—TMSR then TMNSR then TMOR. The designations of June 1 provide an example of this. There was sufficient reserve availability that afternoon to meet the TMSR requirement, but all the inexpensive MWs available were then consumed. Hence, TMNSR and TMOR had to use more expensive resources. Units that normally receive TMNSR or TMOR designations were given TMSR designations instead, so the price had to increase for the inferior reserve bid stacks to clear the markets. One asset was designated as providing 119 MWh of TMSR, 197 MWh of TMNSR, and 8 MWh of TMOR on hour 12 of June 1. The same asset's designations became 368 MWh of TMSR, 12 MWh of TMNSR, and 0.75 MWh of TMOR by hour 14. This bidder's resources were shifted away from the inferior reserves despite a TMNSR bid of $0.65 and a

TMOR bid of $0.55. A smart buyer of ancillary services would have procured TMSR from another source in order to save this bidder for the inferior services. The result of this inefficient procurement was a complete inversion of prices—the TMOR price ($129) exceeded the TMNSR price ($125), which exceeded the TMSR price ($.80). The inversion of reserve prices, which is common to other sequential systems, like California, occurred in 25 hours in June, 17 hours in July, and 8 hours in August. The cap on reserve prices partially corrects this problem, but the basic inefficiency of the sequential market persists.

2.5 Operating procedure 4

Throughout the summer the ISO responded to tight system conditions by implementing Operating Procedure 4 (OP4). This allowed the Real Time SPD to drop the reserves that it cleared below what was required. Over the course of the day on June 7 the actual reserve designations assigned fell below 40% of requirement for TMSR and eventually fell to zero for both TMNSR and TMOR.[7] Similar actions were taken on other dates during the summer including June 8, June 28, July 6, and July 16. When load rose at the end of July, reserve prices rose, but the designations remained high.

Price spikes in the reserve markets created large windfalls for those that received designations, but nonetheless the cost burden of supplying reserves was limited. Reserve prices were consistently high in OP4 hours, but total designations were low, so the product of the two did not rise as drastically as the price. Figure 25 shows that the cost of supplying the three reserves very rarely exceeded 5% of the total payments for energy and reserves. Figure 25 also demonstrates that in the hours that the system experiences the highest load levels the share of payments made for inferior reserves falls[8]. It rarely became necessary for the SPD to drop the amount of TMSR that it cleared below what was required, so in the tightest hours most of the payments for reserves go to TMSR.

The limited cost of supplying reserves should not be declared a great success of the markets. A bidder who knows that OP4 will be implemented has incentive to bid excessively high in both the reserve markets *and* the energy market. If the energy bid is so high that it is not accepted, there is no great loss for the bidder – his reserve bids necessarily will be accepted and a high reserve price will be paid. The interdependency of the markets lets the inefficiency of the reserve markets create distortions in the energy market. In hours when OP4 is implemented, the size of the reserve markets may shrink, but the distortions caused by inefficient reserve prices grow.

2.6 Bidding behavior

Figures 27 and 28 show an estimate of the aggregate bid stack in the TMNSR market for each hour of June 7. The quantity available from each generating unit was calculated as High Operating Limit (HOL) minus generation minus TMSR designation. This somewhat overestimates availability due to ramp rate constraints. Some bidders rarely receive TMNSR designations despite low TMNSR bids. These bidders were not included in the aggregate availability totals. The figures show that the volatility in the reserve market prices is not caused by participants drastically changing their bids from hour to hour. The volatility arises because as generation and superior reserve category designations increase, the bid stacks get shifted dramatically to the left. If the TMNSR stack shifts such that the plateau in the 3 to 4 dollar range is marginal, prices stay low, but as soon as this plateau is passed prices escalate.

---

[7] See Attachment 3 of the June 7–8 Audit Report

[8] Energy revenues for each hour shown in Figure 25 are calculated as Energy Clearing Price times Load Meter Readings plus Uplift Payment plus Congestion Uplift Payment. Since the energy market is a residual market not all of this revenue is received in market transactions, but the value of energy produce to serve native load is best estimated using the clearing price of the pool. The percentages referred to are the percentage of energy revenue plus TMSR revenue plus TMNSR revenue plus TMOR revenue.

The bid stacks move slightly to the right at a price of $150. Two lead participants bid four generating units in the TMNSR market at a price of $150 in every single hour from May 1 to August 31. These units were given positive designations of as high as 30 MWh in 92 different hours between June 1 and August 17. The availability added by the 4 units is minimal (the stacks move only slightly to the right at a price of $150), but nonetheless these units had a large influence on price formation in many hours of June and July. Not surprisingly, both of the lead participants control other units that bid lower prices in the TMNSR market. All of these units benefited when the price was set at $150.

Tables 30, 31, and 32 give some indication of the bidding behavior of market participants for the entire period of May through August. Tables 30 and 31 show that fossil fuel burning generators tend to submit 0 bids into the TMNSR and TMOR markets whereas hydro resources are more likely to submit high bids and set the price. Because hourly prices are averages of five-minute prices, and because more than one unit can submit the same bid, more than one unit can help set the clearing price in any given hour. If a unit submitted a bid that was greater than or equal to the clearing price, and received a positive designation, it is probable that the unit set the clearing price for at least some of the hour. The inter-temporal optimization of the SPD causes some high bids to be accepted even though they do not help set the clearing price, but this is relatively rare. In the majority of hours there is only one bidder with a bid as high as the clearing price.

Some interesting conclusions can be drawn from a comparison of Tables 30 and 31. Table 30 indicates that 39% of all the TMNSR designations that the average hydro non-run of river unit received between May 1 and August 31 came when the unit submitted a positive bid that was at least the clearing price. This can be compared to the corresponding number in Table 31 that says that only 15% of all the TMNSR revenue that the average hydro non-run of river unit received between May 1 and August 31 was received when the unit bid at least the clearing price. The difference between these two numbers indicates that these units tended to set the TMNSR price when the price was low (but positive). In contrast, the average hydro run of river unit received 21% of its designations when it was bidding at least the TMNSR clearing price but received 38% its of revenue during those hours, indicating that the price must have been high during those hours.

Table 32 shows that the largest lead participants received a large proportion of their designations and revenues in hours that they were helping to set the price. A single lead participant can have many different generating units, each with a different TMNSR bid price in any given hour. If any of a lead participant's generating units bid at least the clearing price and was given a positive designation then all of that lead participant's units' TMNSR revenue for that hour was included (likewise for TMOR). For anonymity, the bidders IDs have been renumbered and ordered from the greatest TMNSR revenue to the least for the May – August period. Of all of the TMNSR revenue that the first TMNSR provider received, 46% was earned when the lead participant had a unit with positive designations that was bidding at least the TMNSR price. The second TMNSR provider most often set the price (1006 out of 2952 total hours with a successful bid of at least the TMNSR price), but when it did, the price was low, so the participant did not receive a great deal of revenue in these hours (only 5% of the participant's total).

3   The ISO's short-term remedies

Recognizing the signs of market failure in the reserves and OpCap markets, the ISO introduced two short-term remedies.

ISO Reserve remedy: Cap the reserve prices with the energy price.

On June 27, 1999, the ISO capped the reserve prices with the energy price. Initially, this was done on a five-minute basis. Whenever the clearing reserve price in the five-minute interval was above the five-minute energy price, the reserve price was reduced to the greater of the energy price or 0. This was

11

applied to all three reserve markets (TMSR, TMNSR, and TMOR). Later, in July 1999, the ISO switched to applying the cap on an hourly basis. Whenever the reserve price in the hour was above the hourly energy price, the reserve price was reduced to the energy price. The shift from five-minute to one-hour application of the cap effectively raises the cap slightly, depending on the variability of prices within the hour.[9]

Despite the cap, there are several cases where the reserve prices were above energy price (see Table 26). Many of these cases are the result of a downward revision of the energy price due to an emergency sale. If the revision occurs more than five days after the event, then the ISO does not have the power under the rules to revise the reserve prices. The few remaining cases are simply processing errors. Only recently has the ISO been able to automate the application of the cap. Thus, for most of the four-month period, the cap was applied manually, and some instances where the cap should have been imposed were missed.

ISO OpCap remedy: Cap the OpCap price in OP 4 conditions with five times the average of the three highest hourly clearing prices in the previous thirty days during non-OP 4 conditions.

The cap on the OpCap price in OP 4 conditions was applied retroactively to the beginning of the market. The cap, however, was not binding until particular days in June.

4   Options for medium-term remedies

The ISO's short-term remedies have been an essential and important step in improving the NEPOOL markets. Both remedies involve the use of price caps. Although price caps are inconsistent with competitive markets, their careful application in response to market design flaws is necessary. The ISO, recognizing the dangers of rigid price caps, wisely decided on market-based price caps, allowing the cap on reserves to vary with the energy price and allowing the OpCap cap to vary with the highest OpCap prices in non-emergency situations. Both of these remedies should be continued until a better solution can be implemented.

The ISO and NEPOOL should continue to work aggressively on a long-term fix to the basic design flaws in the reserves. The recent work of NEPOOL's Congestion Management and Multi-settlement Committee is an important step in the right direction. However, realistically it is unlikely that the long-term solution will be implemented until 2001, given the complexity of the issues involved. Hence, it is important to identify a medium-term fix that the ISO can implement within a matter of months. I present such a medium-term solution below.

Recommendation: Eliminate the OpCap market.

The market flaws in the OpCap are severe. Fortunately, there is an easy medium-term fix that is also a long-term fix—eliminate the OpCap market. One cannot argue that the OpCap market serves any useful purpose as presently designed. Since its operation has real costs with no apparent gain, the market should be eliminated. Capacity should be rewarded in the energy and reserve markets, and not in some phony market.[10]

---

[9] If within hour prices are fairly stable, this change has little effect; if within hour changes are large, the effective raise in the cap is large.

[10] If participants want to play the "ask and it shall be given" game, they should do so on a voluntary basis. Moreover, the activity should not be orchestrated by the ISO, since it has nothing to do with running a reliable and efficient electricity market. I doubt that the game would survive such a market test.

Recommendation: Adopt the smart buyer model for the reserve markets.

I recommend that the reserve markets continue in the medium-term, but with the ISO adopting a smart buyer model for reserves. As a smart buyer, the ISO:

1. Never pays for additional reserves more than the economic value of the additional reserves.

2. Reduces its demand for reserves as reserve prices increase.

3. Shifts purchases toward higher quality reserves when they are priced less than lower quality reserves.

In conducting the markets for reserves, the ISO is effectively purchasing reliability on behalf of load. The ISO has a responsibility to purchase wisely, given the aggregate preferences of load. On this basis, the ISO as smart buyer develops a demand curve for reserves that reflects the marginal value that additional reserves have for the system.

The essential elements of the smart buyer model can be adopted in the medium-term. The ISO recently has implemented point 3 above. The reserves now are treated as a cascade with the ISO filling its need for lower quality reserves with higher quality reserves if the higher quality can be purchased at a lower price. The difficult step is establishing a demand curve for reserves, which reflects the marginal value of additional reserves. This will accomplish points 1 and 2 above. It will require some discussion among participants and regulators, but once new operating procedures are established the ISO can begin purchasing reserves according to its demand curve, rather than the current vertical "demand" curve, which does not reflect the marginal value of additional reserves.

The demand curve for reserves may sound like an arbitrary object. One may fear that determining the demand curve will quickly turn into heated debates about the value of lost load. Although I anticipate lively discussions, I believe that a reasonable approximation of the demand curve can be constructed in an implementable way. The key is focusing on what a demand curve is. The demand curve for reserves specifies for every quantity of reserves the marginal value of additional reserves. Today there are tractable methods for determining these values from the shadow prices of the appropriate optimization problem.

In what follows, I will assume that the ISO is able to construct a downward-sloping demand curve for reserves.

Recommendation: Restructure the reserve markets.

The basic flaws in the reserve markets stem from two features of the current markets:

3. The vertical "demand" curve, derived from a rigid reserve requirement, implies that prices are arbitrarily high in times of scarcity.

4. Losing bidders face the same obligations as winning bidders.

The first flaw is eliminated by constructing the true demand curve for reserves, based on the marginal value of additional reserves. The second flaw can be solved in two ways.

The first way is to construct one or more forward reserve markets. Then the losing bidders would learn that they are not needed to provide reserves, while they still have time to find something to do with their excess supply, such as offering it in another market. This is the approach used in California and some other markets (Chao and Wilson 1999). I do not believe that this is a feasible solution in the medium-term. The construction of new markets cannot be done in a few months. Day-ahead reserve markets are planned in the long run as part of the CMS/MSS Straw Proposal.

The second solution to flaw 2 is to revise the structure of the reserve markets in a way that is consistent with the basic market features in place today. These basic features are (1) ex post clearing, and (2) an obligation of all operable capacity to participate in the market (i.e., respond to the dispatch instructions of the ISO). Features (1) and (2) imply that the true supply curve is in fact vertical—a fixed supply of reserves is offered to the market regardless of price. This is illustrated in the diagram below. The first pane illustrates the current market structure for reserves. The bidders submit offers from which the aggregate "supply" curve (S) is formed. The ISO establishes the reserve requirement, forming the vertical "demand" curve (D). The ISO then designates the reserves (Q_D) and the clearing price (P). Those with bids below P are paid; those with bids above P are not. All available supply (Q_S) provides reserves in the sense that they provide dispatch flexibility to the extent that they can respond to dispatch instructions.

Under the revised market structure illustrated in pane (b), the clearing price is found by intersecting the true supply curve (S_T), computed by the ISO based on the participants' ability to respond to dispatch instructions, and the true demand curve, computed by the ISO from its dispatch and ancillary optimizations. The result is the price (P_T), which is received by every participant that is providing needed dispatch flexibility.
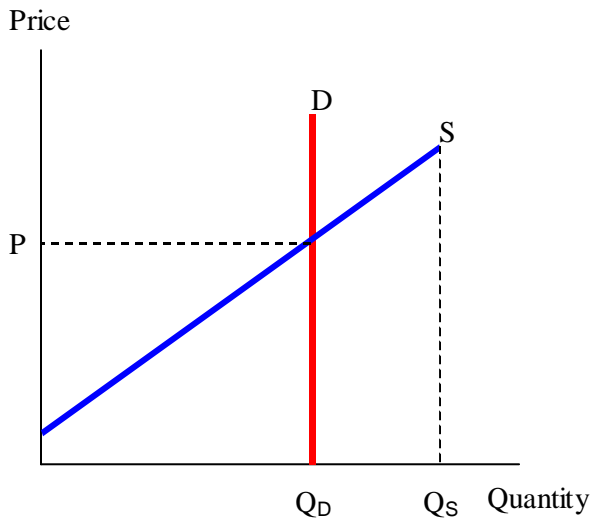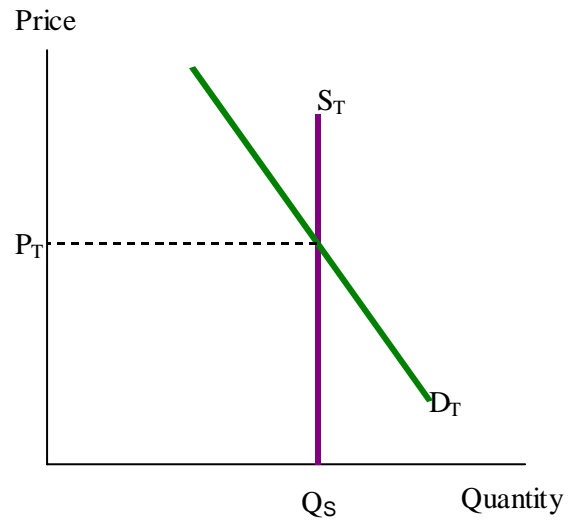
(a) Current Market Structure for Reserves   (b) Revised Market Structure for Reserves

The revised market structure replaces the flawed markets for reserves with markets based on sound economic principals.

5   Downward-sloping demand for reserves

Under the revised market structure, rather than formulating three inflexible reserve targets, the ISO creates three demand curves for reserves that reflect more accurately the true marginal value of each additional MW of resource. This marginal value can be identified using probabilistic reserve valuation techniques. The demand curve depends on the state of the system and the forecast load. Since the probability that additional reserves will be useful decreases with the total amount of reserves available, demand curves for reserves are downward sloping.[11] The specific slope and shape of these curves depends

---

[11] Typically, the marginal value of reserves is determined from the product of the probability of lost load and the value of lost load: POLL(t)·VOLL(t). However, the calculation may also include factors such as the probability of incurring a NERC penalty for maintaining inadequate reserves, the size of the penalty, and transmission risk.

on the potential impact on reliability. As shown in pane (c), the demand curve for spin intersects the vertical axis at P_{max}. This may be the value of lost load (VOLL), since without any spin the ISO must surely shed load. However, it is also possible for P_{max} to reflect other factors as described in the prior footnote. Alternatively, in the short and medium term, it may make sense to set P_{max} equal to the greater of 0 and the energy price. This is an especially sensible option if it is viewed that there are structural problems in the reserve markets at certain times.

For notational convenience, I refer to the TMSR market as spin, the TMNSR as 10, and the TMOR as 30. Thus, P_{spin}, P_{10}, and P_{30} are the prices in the TMSR, TMNSR, and TMOR markets, respectively. These prices are determined by the intersection of the appropriate demand curve with the supply that is available in real time. I consider two cases. The first, depicted in panes (c)-(e), is the more typical case when reserves are plentiful, and hence reserve prices are low. The second case, depicted in panes (f)-(h), shows a situation when reserves are more scarce.

The demand curve for spin in pane (c) remains fairly flat up until ½ of the first contingency, since the largest contingency alone would require the shedding of load when 10 minute reserves (spin plus 10) are less than this first contingency. After this point the demand curve falls fairly steeply, since shedding of load is only required in the face of multiple contingencies. Since the contingencies are largely independent, the chance of multiple contingencies is much less than the chance of a single contingency. Eventually, the demand curve converges to zero, but it always remains above zero, since it is always possible—regardless of the size of reserves—for sufficient generation and transmission to fail such that load will need to be shed.

Prices and quantities for the three reserve products are determined sequentially: spin, ten-minute non-spin, and then thirty-minute operating reserve. Pane (c) shows the market for spin. The relatively large quantity of spin available,[12] Q_{spin}, means that the supply and demand curves intersect at a relatively low price P_{spin}.[13] Pane (d) shows the market for ten-minute reserves. This market takes as given the quantity Q_{spin} that was designated for spin. The price of ten-minute non-spin, P_{10}, is determined from the intersection of the demand curve for spin+10 and the quantity of spin+10 that is available. Since spin can always substitute for non-spin, P_{10} is necessary no more than P_{spin}. Finally, pane (e) shows the market for thirty-minute reserves. Again, the quantity of spin+10 is taken as given, and the demand curve begins no higher than P_{10}, since spin or 10 can substitute for 30.

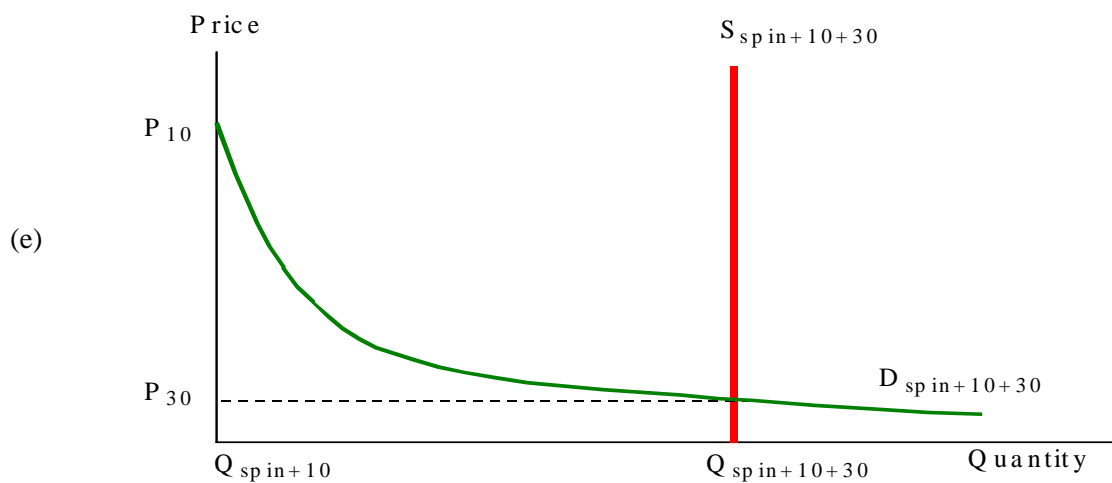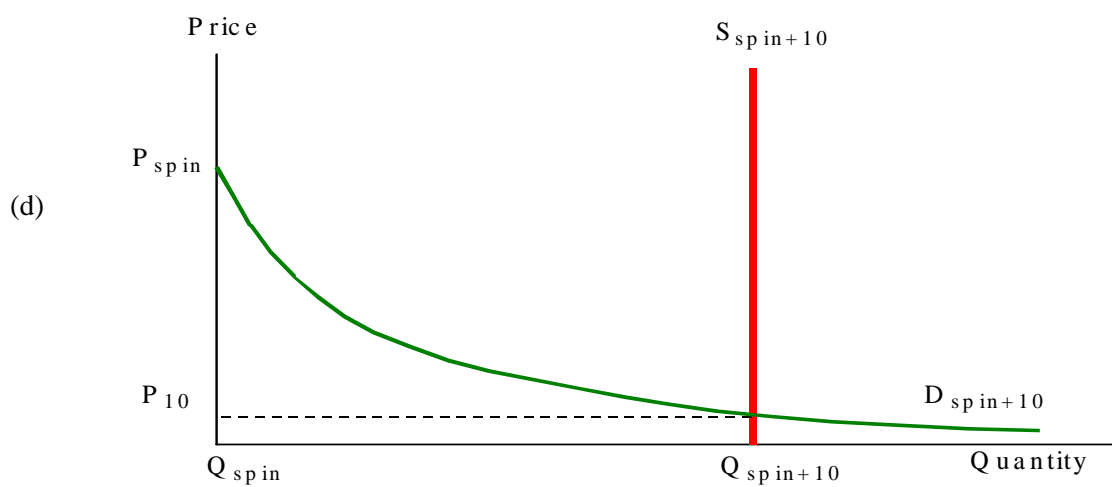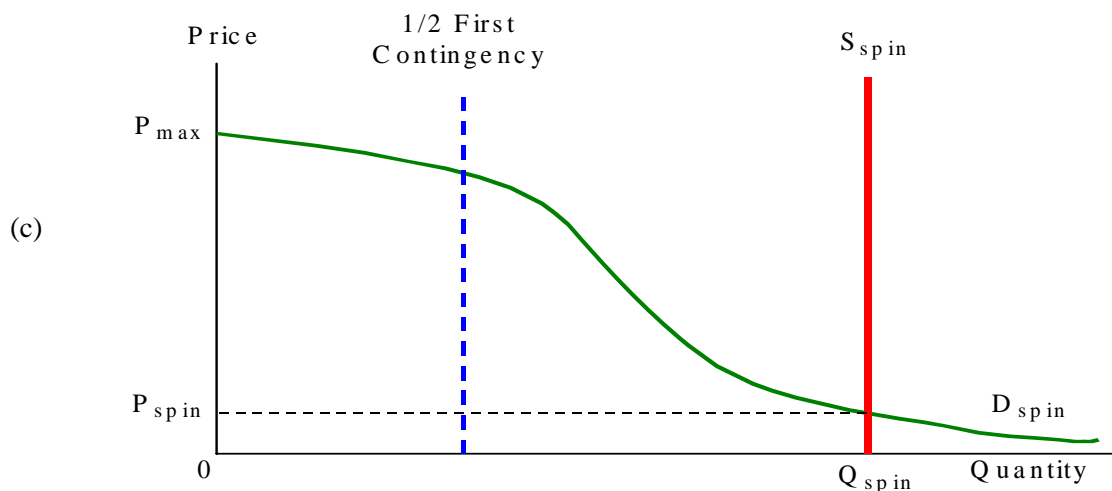Panes (f)-(h) show analogous figures for the three markets when reserves are more scarce.

---

[12] Both hydro and pump storage should be treated as spin, since both services provide a service that is at least as good as spin, even when they are not online.

[13] Opportunity cost measures derived from energy bids are important in deciding which online capacity is called for energy and which is left over to provide spinning reserves, but there is no need for this opportunity cost to determine the spin clearing price or for it to determine which units supplying spin get paid for it. Unit commitment should reserve a level of spin such that the spin price is high enough to cover the lost opportunity cost of all units. Hence, a floor on the spin price is the highest lost opportunity cost of any unit backed down to provide spin.
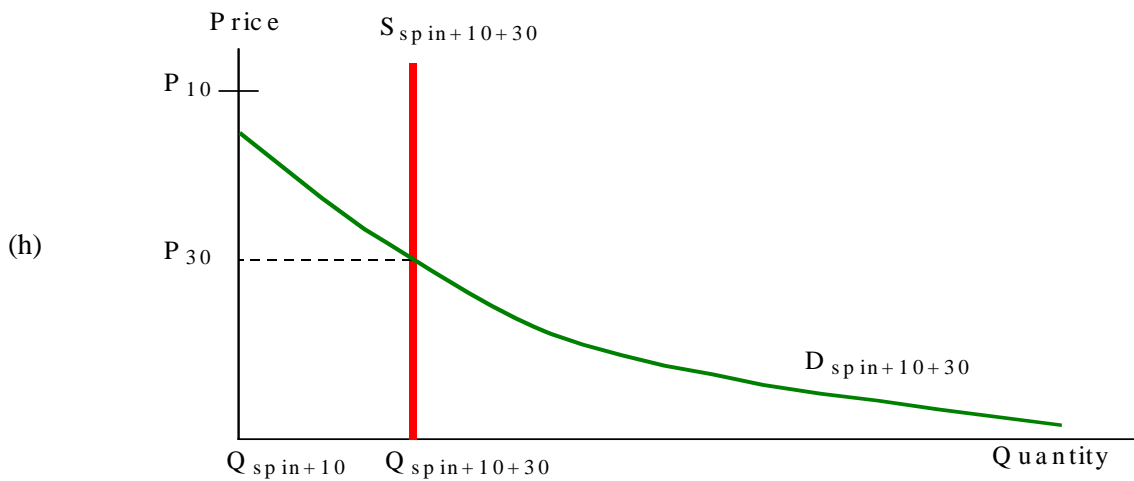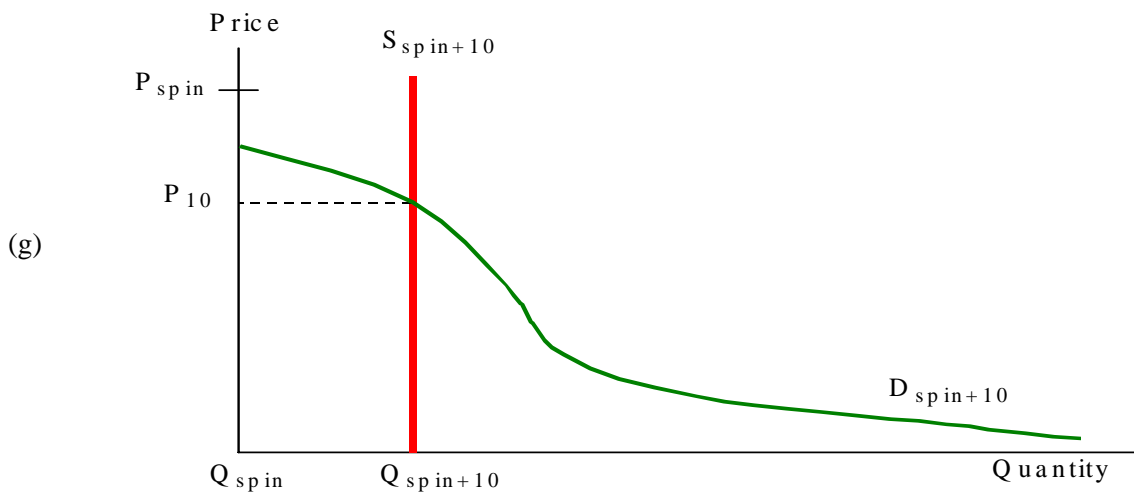
Reserve Market in Typical Case (Surplus)

Spin = TMSR   10 = TMNSR   30 = TMOR

(c)

(d)

(e)

Reserve Market when Supply is Tight

Spin = TMSR   10 = TMNSR   30 = TMOR

(f)

Price

1/2 First Contingency

S_{spin}

P_{max}

P_{spin}

0

Q_{spin}

D_{spin}

Quantity

(g)

Price

$S_{spin+10}$

P_{spin}

P_{10}

Q_{spin} $Q_{spin+10}$

$D_{spin+10}$

Quantity

(h)

Price

$S_{spin+10+30}$

P_{10}

P_{30}

$Q_{spin+10}$ $Q_{spin+10+30}$

$D_{spin+10+30}$

Quantity
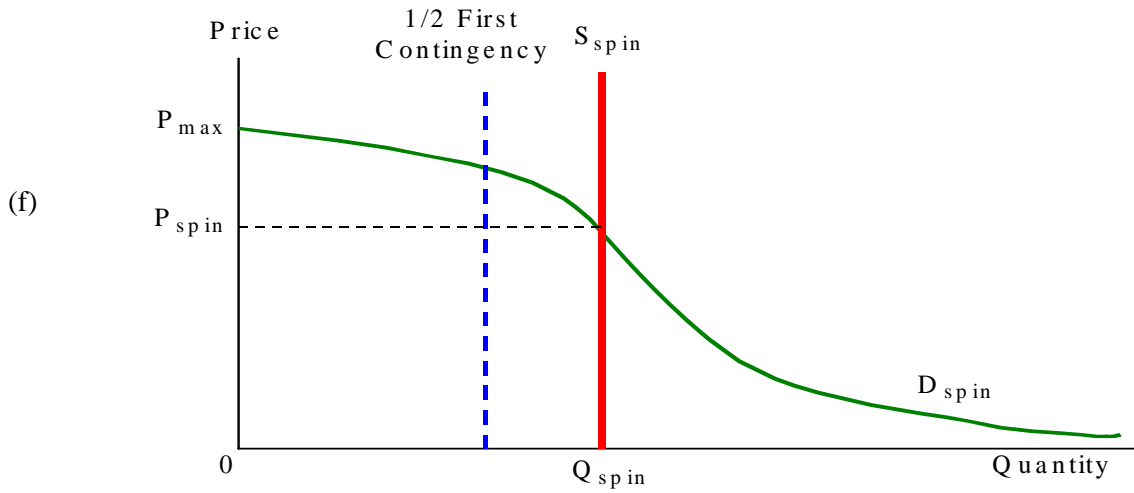
5.1 The marginal value of reserves

The purpose of the reserve markets must be identified before any technique for the estimation of marginal value can be determined. Ideally real time reserves should be used solely to protect the reliability of the system in the event of generator forced outages (Hirst and Kirby 1998). However, in New England the lack of day-ahead markets for operating reserves shifts some of the burden of protecting against load-forecast errors to the real time reserves. An estimation of the true marginal value of additional reserves must include the contribution to bulk-power reliability as well as the commercial value added by the additional system capacity.

Increased levels of operating reserves enhance the ability of a system to quickly return to pre-contingency levels of frequency and area-control error (ACE) after a generator outage, however, little work has been done to quantify this relationship (Hirst and Kirby 1998). In addition, the cost of extended frequency disruptions needs to be researched in order to rationally develop criteria for maximum allowable recovery time. The judgment of control-area operators and engineers may have been sufficient to develop reliability standards in the past, but full specifications of the cost of reliability will be necessary to promote the efficiency of competitive markets.

Significant theoretical work has been done on the valuation of scheduled reserve capacity, particularly for spinning reserve. Sufficient scheduled reserve capacity can correct for load forecasting errors and generator outages without load shedding. The value of reserves can be calculated as the value of the expected load that would have been shed if reserves were not available. It should be pointed out that the engineering literature has focused on scheduled reserves, not reserves that are designated in real time. Scheduled reserves can be thought of as call options. There is some positive probability that the system operator will exercise the option and use the reserve for energy in real time. When reserves are designated in real time there is zero probability that the reserve is used for energy; a unit is designated as providing reserves or energy, not both. From the perspective of real time the only relevant uncertainty pertaining to reserves is the possibility of a contingency arising in the very short-term future. However, there is still a great deal of uncertainty about load and outages when generators are deciding to make capacity available, and it is this decision that is effected by the price signals sent by the reserve markets. Therefore, when market designers derive probabilities of contingencies occurring in real time for the purpose of reserve valuation, the probabilities should be taken from the perspective of a generator making an availability decision. If it is assumed that generators make their availability decisions one day in advance, then the probabilities of contingencies used to value real time reserves are identical to the probabilities of contingencies used to value reserves scheduled day-ahead.

Use of probabilistic techniques will capture the random nature of reliability better than deterministic criteria. The uncertainty in load is more important than its actual level in determining optimal levels of reserves; as the distribution of load is more variable reserves are more valuable. Reserve requirements are often tied to the capacity of the largest scheduled generator. If the largest generator is very reliable there is no reason for it to have much influence on reserve requirements. Incorporation of the reliability of all scheduled units leads to a better description of system conditions.

The valuation of scheduled reserves begins with the development of the capacity outage distribution function for every possible state of the system. This function should incorporate the reliability of all scheduled generators as well as load conditions (Billinton and Fotuhi-Firuzabad 1996; Billinton and Karki 1999; Gooi et al. 1998). An estimate of the number of MWhs that would have to be shed in order to maintain system reliability in every contingency can be convoluted with the capacity outage distribution function to derive the expected energy not supplied.[14] The product of the expected energy not supplied

---

[14] Multiply the load that must be shed in each contingency by the probability of the contingency given the state of the system, and sum all the resulting products.

and the value of lost load (VOLL) gives the total expected cost of energy not supplied. The decrease in this total cost as reserve levels change by one MW determines the marginal value of reserves. This calculation can be done for every reserve level for every possible state of the system thus deriving a series of marginal value curves. The rules for mapping system conditions into demand curves should be determined before operations. To simplify calculations Stremel et al. (1980) show that the equivalent load curve (load plus outages) can be represented using cumulants.[15] Billinton and Karki (1999) show that the use of a Monte Carlo simulation can also simplify calculations and allow for the inclusion of system effects that are difficult to include in a direct analytical approach.

Gooi et al. (1998) develop a method for incorporating the estimated value of spinning reserves into a Lagrangian Relaxation Unit Commitment program. Overall cost savings are achieved when the reserve requirement is adjusted based on the balance of costs and benefits during unit commitment (Tseng et al. 1999; Guan and Luh 1996).

The value of lost load is a crucial unobservable parameter in the valuation of reserves. Reserve prices should never exceed VOLL since load can always be shed to produce reserves. Estimates of VOLL range from $2/kWh to $25/kWh. Surveys show that VOLL is dependent on the duration of interruptions, the number of interruptions, the time of day, the time of year, availability of advance warning, the location of the interruption, as well as other variable factors. If VOLL is assumed to be constant, and is set to the average of all of its possible values, it will not precisely reflect the true cost of shedding load at any given time (Kariuki and Allan 1996). The estimation of VOLL needs to be central to the reserve valuation debate. The precision of algorithms used for reserve valuation will be highly dependent on the accuracy of VOLL estimates.

5.2 Lessons from the English experience

The central energy market in England and Wales, known as the pool, has used capacity charges and availability payments since its inception. These payments are determined day-ahead based on expected reserve margins and so are fundamentally different from what is proposed here. However, these payments demonstrate the potential problems that arise when payments are made to suppliers based on the total availability of reserves.

The capacity charge is part of the day-ahead pool purchase price and is paid per MWh of load served. The charge is calculated by subtracting the System Marginal Price (the price that clears the energy bid stack) from the estimated value of lost load (VOLL) and then multiplying by the loss of load probability (LOLP). VOLL was originally set at £2 per kWh and has increased annually. The LOLP is calculated for each half hour period of the coming day using conventional probability theory, and is a decreasing function of the expected amount of excess capacity available. Generators that have capacity available but not actually used receive the availability payment. A unit's payment is calculated by subtracting the higher of the System Marginal Price and the unit's bid price from VOLL and multiplying this by LOLP.

The capacity charge and availability payment were designed to reward suppliers who were available when high load or generator outages cause the system to become tight. Economists criticized these market rules because suppliers can determine when the system becomes tight by strategically timing planned outages. Since LOLP increases as expected excess capacity decreases generators are able to gain large payments by withholding capacity. Bunn and Larsen (1992) show that the extreme convexity of LOLP at low levels of the expected reserve margin makes the incentive to withhold capacity especially strong when the expected reserve margin is low and the system is most vulnerable. Wolak and Patrick (1998)

---

[15] Cumulants are linear combinations of statistical moments.

demonstrate that the two largest participants in the pool tend to reduce their availability below the levels that would be predicted by NERC average availability factors, and tend to reduce availability more often than other participants. This behavior is consistent with strategic capacity withholding because it is the largest participants that have the most to gain from reducing their availability. Wolak and Patrick also find that capacity withholding leads to low reserve margins even in hours where the System Load is relatively low.

Like the English rules, the proposed revisions to New England's reserve markets pay high amounts when availability is low. It is crucial that the markets be designed so that availability is not made low strategically in order to force high payments. For the following reasons I believe this is possible:

1. The proposed rules do not involve anything analogous to the English capacity charge. Suppliers of energy are rewarded with high prices in peak hours because the market clears far up the energy bid stack. There is no reason to add an ad hoc payment to the energy price reflecting the tightness of the reserve markets. The incentive to withhold capacity is obviously much less when the reserve margin only effects payments made to units providing reserves. Since reserve totals are a small fraction of energy produced, doing away with the capacity charge does away with most of the incentive to withhold capacity.

2. Market power in England and Wales is more of a problem than in New England. Deregulation in England and Wales resulted in only two firms that regularly set the price. In New England the two largest Lead Participants account for approximately half of the capability and slightly more than half of the revenue in the reserve markets.

3. Demand curves can be derived with sufficient elasticity so that a reduction in availability does not lead to an excessive price increase. Wolak and Patrick make a similar observation about the English market: "The true value of the VOLL to consumers and the true relationship between the estimated reserve margin and LOLP are unknown to the regulator, but the magnitude of VOLL or the function specified to relate the expected reserve margin to the LOLP can have a tremendous impact on the observed market clearing prices. For example, setting the VOLL too high can make the payoff from strategic capacity choice timed with demand fluctuations during days and across the year a profitable strategy. If the relationship between the LOLP and the expected reserve margin is too steep in absolute value, this can also increase the profitability of this strategy. Consequently, the regulator overseeing the operation of the pool should view such variables as the VOLL and the function relating the reserve margin to the LOLP as instruments for obtaining the desired market outcomes, rather than as fixed constants or relationships." The ISO should not be constrained to choosing demand curves that reflect the marginal benefits of additional reserves precisely. If the marginal value is calculated to be very steep over a range of reserve levels, a mechanism can be used to flatten out the demand curve in that range. For example, the algorithm used to derive demand curves can be constrained to keep the demand curves above a specified elasticity level.

4. A well-defined algorithm for determining each demand curve should be implemented, and its form communicated to all participants, but the ISO must retain the power to adjust the parameters and methodology of the algorithm in response to market performance. It is important that suppliers of reserves understand the market structure in order to participate efficiently, however, this should not constrain the ISO's ability to make quick and decisive corrections. Corrections of the market flaws in England and Wales have been slow in coming.

5.3   *Reliance on only real time reserve markets should be temporary*

The proposed revisions should be considered as temporary corrections. The value of scheduled reserves can be clearly defined by their contribution to the probability that load will be met. The value of

real time reserves is harder to define since capacity is only designated as providing reserves if it does not serve load. Operating reserves protect the system reliability of the short-term future, but by definition provide no tangible service in the present. Because clear definitions of value can be derived for scheduled reserves, clear price signals can be sent leading to efficient incentives.

6   The reserve markets in the long-term

For the reserve markets to be efficient, it will be necessary to create one or more forward reserve markets. Then the losing bidders would learn that they are not needed to provide reserves, while they still have time to find something to do with their excess supply, such as offering it in another market. Chao and Wilson (1999) describe one approach, which has been adopted in California. Although construction of a new market cannot occur in a few months, the ISO and NEPOOL participants should continue their efforts toward a long-term solution of the reserve markets.

The development of a long-term solution is beyond the scope of this paper. However, I will comment briefly on the design effort to date. The CMS/MSS Straw Proposal of November 1999 presents a multi-settlement proposal for reserves. In the day-ahead market, reserves are scheduled and generators make financial commitments to provide reserves. The real time market serves a balancing function. Deviations from the day-ahead commitments are settled at the real time prices. The real time reserve markets are structured just as I have described here. The smart buyer model is employed both day-ahead and in real time. Thus, the CMS/MSS Straw Proposal is fully consistent with all of the recommendations made here. This will greatly ease the transition from the medium-term to long-term. In addition, the costs to implement the medium-term recommendations are easily justified, since all the tasks are required to implement the long run solution. In this sense, the medium-term solution is a natural step toward the long-run solution.

One issue that has not yet been addressed in the reserve market discussions is the idea of decremental reserves. At times, the balancing problem is that there is too much generation. The reserve market should provide downward as well as upward flexibility. Doing so may have a positive impact in the market for incremental reserves. One of the major problems of the real time reserve markets is that they are inherently thin. One way to mitigate this problem is to rely more on forward markets. Another way is to allow for greater arbitrage opportunities between energy and reserve markets. By including all four combinations of demand/supply and incremental/decremental bids in the energy and reserve markets, new arbitrage possibilities will be created in these markets. As a result, the reserve markets will be more liquid and less prone to insufficient bids.

7   Conclusion

The OpCap and reserve markets have serious flaws that must be addressed. The ISO's short-term fixes have been necessary and effective at addressing the immediate problems. However, better solutions can be adopted in the medium term. In particular, I recommend:

1. Eliminate the OpCap market.

2. Establish a downward sloping demand curve for reserves. The demand curve would be capped in real-time by the energy price or zero, whichever is larger.

3. Pay the clearing price to all resources that provide the service.

4. Establish the true real-time supply curve as simply the quantity of the resource made available in real time. The current service availability bidding would be eliminated.

5. Establish back down bids in the TMSR market. Bids would be infrequent, perhaps monthly.

6. Never set a price in the TMSR market less than the largest lost opportunity cost.

7. Continue to cascade the quantities of the bids between operating reserve products. (Bid prices would not be cascaded because they would not exist.)

8. Correct the classification of off-line units that provide a service that looks and acts like TMSR.

All of these changes are consistent with the CMS/MSS Straw Proposal. These changes represent an important step toward the long-term solution involving multi-settlement energy and reserve markets. These markets should be designed carefully to address the basic economic and engineering issues necessary for an efficient wholesale electricity market.

Appendix: Overview of the markets

ISO New England conducts seven interdependent markets in its operation of NEPOOL's wholesale electricity market: (1) the energy market, (2-5) four markets for ancillary services, and (6-7) two capacity markets. A description of each is given below. For brevity, many important details are omitted.

1. The *energy market* is a residual market. Only the difference between a participant's energy resources and its energy obligations is traded in the ISO market. These resources and obligations include amounts covered by bilateral contracts. Hourly bids, expressed in $/MWh, are submitted on a day-ahead basis for the next 24 hours. The ISO then schedules the generating units that will run the following day based on minimizing total costs in the energy market, as represented by the accepted bids. The market is settled after the fact on an hourly basis. All transactions are priced at the (ex post) energy clearing price. Payments/receipts are equal to the MWh bought/sold times the clearing price. Suppliers are paid for out-of-merit-order dispatch to alleviate transmission congestion on the basis of their bids submitted in the energy market.[16]

2. The *ten-minute spinning reserve (TMSR) market* is a full requirements market. All TMSR is bought/sold through the ISO. Bidding and settlement are done as in the energy market—hourly bids in $/MW for the next day are submitted, and the markets are settled hourly after the fact. Given the units dispatched to provide energy, the ISO selects the least-cost resources to provide required TMSR, taking into account bids and lost opportunity costs. Designated resources are paid the energy clearing price for any MWh provided. In addition, for every MW designated as spin, resources are paid the clearing price for spin, which is calculated from the bid plus lost opportunity cost. The total cost of providing TMSR is shared proportionally by load.

3. The *ten-minute non-spinning reserve (TMNSR) market* is a full requirements market. All TMNSR is bought/sold through the ISO. Bidding and settlement are done as in the energy market—hourly bids in $/MW for the next day are submitted, and the markets are settled hourly after the fact. Designated resources are paid the clearing price times the MW provided as reserved capacity. The total cost of providing TMNSR is shared proportionally by load.

4. The *thirty-minute operating reserve (TMOR) market* is a full requirements market. All TMOR is bought/sold through the ISO. Bidding and settlement are done as in the energy market—hourly bids in $/MW for the next day are submitted, and the markets are settled hourly after the fact. Designated resources are paid the clearing price times the MW provided. The total cost of providing TMOR is shared proportionally by load.

---

[16] The bidder may not be paid its bid if certain transmission congestion structure and price screens are not satisfied.

5. The *automatic generation control (AGC) market* is a full requirements market. All AGC is bought/sold through the ISO. Bidding and settlement are done as in the energy market—hourly bids for the next day are submitted, and the markets are settled hourly after the fact. AGC is measured in *regs*, which measures a unit's ability to follow load. Units that can provide AGC at lowest cost based on bids, lost opportunity costs, and production cost changes are selected. Generators providing AGC are paid the clearing price for time on AGC times the number of regs *plus* a payment for AGC service actually provided *plus* any lost opportunity cost. The total cost of providing AGC is shared proportionally by load.

6. The *operable capability (OpCap) market* is a residual market. Only the difference between a participant's operable capability resources and its operating capability obligation (load plus operating reserve) is traded through the ISO. Bidding and settlement are done as in the energy market—hourly bids in $/MW for the next day are submitted, and the markets are settled hourly after the fact. A clearing price is calculated based on the bids of those participants with excess operable capacity. Participants who are deficient in operable capability pay the clearing price for each MW to those who are in surplus and who bid a price less than or equal to the clearing price.

7. The *installed capability market* is a residual market. Only the difference between a participant's installed capability resources and its installed capability obligation (load plus installed operating reserve) is traded through the ISO. Trading in this market occurs monthly. Bids are submitted in $/MW-month on the last day before the month begins. A clearing price is calculated based on the bids of those participants with excess installed capacity. Participants who are deficient in installed capability pay the clearing price for each MW-month to those who are in surplus and who bid a price less than or equal to the clearing price. The market is settled ex post.

References

Billinton, Roy and M. Fotuhi-Firuzabad (1996), "Reserve Capacity Assessment in Small Isolated Electric Power Generating Systems," *IEE Power Engineering Journal*, 10, 73-80.

Billinton, Roy and Rajesh Karki (1999), "Application of Monte Carlo Simulation to Generating System Well-being Analysis," *IEEE Transactions of Power Systems," 14, 1172-1177.*

Bunn, Derek W. and Erik R. Larsen (1992), "Sensitivity of Reserve Margin to Factors Influencing Investment Behaviour in the Electricity Market of England and Wales," Energy Policy, 420-429.

Chao, Hung-po and Robert Wison (1999), "Multi-Dimensional Procurement Auctions for Power Reserves: Incentive-Compatible Evaluation and Settlement Rules," Working Paper, Stanford University.

Cramton, Peter and Robert Wilson (1998), "A Review of ISO New England's Proposed Market Rules," White Paper, Market Design Inc., September 1998.

Gooi, H. B., D. P. Mendes, K. R. W. Bell and D. S. Kirschen (1999), "Optimal Scheduling of Spinning Reserve," *IEEE Transactions on Power Systems*, forthcoming.

Guan, X. and Peter B. Luh (1996), "Power System Scheduling with Fuzzy Reserve Requirements," 11, 864-869.

Hirst, Eric and Brendan Kirby (1998), "Operating Reserves and Bulk-Power Reliability," *Energy*, 23, 949-959.

Kariuki, K. K. and R. N. Allan (1996), "Evaluation of Reliability Worth and Value of Lost Load," *IEE Proc.-Gener. Transm. Distrib.*, 143, 171-180.

Joint NEPOOL CMS/MSS Committee (1999), "CMS/MSS Straw Proposal," working draft, October 14, 1999.

NEPOOL Multi-Settlement Committee (1999), "Multi-Settlement Proposal," Final Report, March 5, 1999.

Singh, Harry and Alex Papalexopoulos (1999), "Competitive Procurement of Ancillary Services by an Independent System Operator," *IEE Transactions on Power Systems*, 14, 498-504.

Stremel, J. P., R. T. Jenkins, R. A. Babb and W. D. Bayless (1980), "Production Costing Using the Cumulant Method of Representing the Equivalent Load Curve," *IEEE Transactions on Power Apparatus and Systems*, PAS-99, 1947-1956.

Tseng, Chung-Li, Shmuel S. Oren, Alva J. Svoboda and Raymond B. Johnson (1999), "Price-based Adaptive Spinning Reserve Requirements in Power System Scheduling," *Electircal Power and Energy Systems*, 21, 137-145.

Wacker, Garry and Roy Billinton (1989), "Customer Cost of Electric Service Interruptions," *Proceedings of the IEEE*, 77, 919-930.

Wolak, Frank A. and Robert H. Patrick (1997), "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market," Working Paper, Stanford University.