A REVIEW OF ISSUES CONCERNING ELECTRIC POWER CAPACITY MARKETS

Project Report

Submitted to
The Maryland Power Plant Research Program
Maryland Department of Natural Resources

Benjamin F. Hobbs, bhobbs@jhu.edu
Javier Iñón, javo@jhu.edu
Department of Geography and Environmental Engineering
The Johns Hopkins University
Baltimore, MD
(410) 516-4681

Matthew Kahal, kahal@exeterassociates.com
Exeter Associates
Silver Springs, MD

December 2001

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This work was supported by the Maryland Power Plant Research Program, with partial assistance for the analysis in Chapter 4 from the National Science Foundation Grant ECS 00-80577. Partial summaries of this work reported here have been published in Hobbs, Iñón, and Kahal (2001a) and Hobbs, Iñón, and Stoft (2001b). We thank Diane Brown and Sandi Patty of the Maryland Power Plant Research Program for their guidance and support. Many people provided us with useful comments, including James Kirby, Bob Hinkel, and Bill Whitehead of PJM; Steve Stoft of the University of California Energy Institute; David Mead and Udi Helman of FERC; Robert Gramlich, formerly of the PJM Market Monitoring Unit; Kent Davis of BGE; and Pete Lalor of Commonwealth Power Corporation. However, the authors are solely responsible for any opinions or errors in this document, which do not necessarily represent the positions of the funding agencies.

ABSTRACT

Restructuring of the electric power industry has been motivated by the desire to increase the efficiency of a sector changed by technological development over the last decades. To achieve this goal, the design of market rules has to account for the special physical and commercial characteristics of electricity consumption, and how society values energy reliability.

This report analyzes alternative proposals for generation capacity markets in electric power markets. Capacity market proposals are a supply-side approach to correcting perceived failures of the market that could result in too little incentive for maintaining existing generation plants and constructing new capacity. The alleged failures include a disconnect between the real-time wholesale energy prices and consumers' energy use decisions; market power; and the possible failure of energy prices to fully capture the system-wide reliability benefits of capacity. A debate is ongoing over whether these problems imply that separate capacity requirement or markets are desirable, and if they are, how they should be designed. The basic questions addressed in this report are whether, in theory, the alternative capacity market designs can yield the right amount of capacity and reliability (allocative efficiency) and produce energy as efficiently as possible (productive efficiency).

A wide range of proposals has been made for capacity markets, each being some variant or combination of two basic approaches: capacity payments and capacity requirements. Each proposal is described, along with a summary of their advantages and disadvantages. Particular attention is paid to the Pennsylvania-Jersey-Maryland (PJM) Interconnection Installed Capacity (ICAP) requirement and market. The ICAP market represents a regulatory approach to improve allocative efficiency, and a market-based approach to achieve productive efficiency, under the assumption that energy spot prices alone would motivate too little capacity investment.

In order to further compare the efficiency of capacity market proposals, we present a stylized model of capacity decision making for a PJM-sized market. This model simulates how competitive providers of peak and baseload capacity react to the incentives provided by alternative approaches to dealing with the capacity question: energy spot markets alone, fixed capacity payments, ICAP-type markets, and reliance on operating reserves markets. The analysis shows that, in theory, the different approaches can each be designed to provide the same amount of capacity adequacy along with the same mix of peaking and baseload capacity. It also illustrates how caps upon prices of energy and operating reserves can prevent attainment of socially optimal levels of capacity in the absence of separate payments or markets for installed capacity.

The report concludes by presenting some implications of these results for PJM. We also identify several additional issues that should be addressed in a quantitative way to provide more definitive conclusions on the relative efficiency of capacity market alternatives.

SUMMARY

The task of market design can be viewed as the creation of a set of rules and incentives so that:

- the socially optimal amounts of the good are produced and allocated to the most valued uses (allocative efficiency); and
- the good is produced by the most efficient means available (productive efficiency).

Concerns of equity and administrative efficiency are also important in market design. The purpose of this report is to analyze alternative proposals for capacity market design in electric power markets. The basic questions addressed are: in theory, do the alternative designs result in the right amount of capacity and reliability (allocative efficiency) and do they produce it as efficiently as possible (productive efficiency)?

In the Pennsylvania-Jersey-Maryland (PJM) Interconnection, an Installed Capacity (ICAP) requirement and market have been created. The ICAP market uses rules to determine the total amount of capacity that is to be installed, but allows the market (through tradable ICAP credits) to determine how that capacity is provided. Thus, the ICAP market represents a regulatory approach to deciding the amount of capacity (allocative efficiency), and a market-based approach to determining who provides it and how (productive efficiency).

However, a debate is ongoing over whether separate capacity requirement and markets are desirable, and if they are, how they should be designed. In theory, a well-functioning spot markets for energy should be sufficient to motivate the optimal amount of investment in generating capacity. However, this assumes that consumer willingness to pay (WTP) for reliability is accurately reflected in price spikes that occur during times of capacity shortage. In reality, however, there are technical and institutional reasons why consumers neither face nor respond to spot prices. Further, because energy prices are often capped, the incentive for capacity additions potentially is distorted, and too little (or, possibly in some cases, too much) capacity might be forthcoming.

Under the assumption that spot prices alone would motivate too little capacity investment, a variety of proposals have been made to correct these market failings. Some address the market failures directly by attempting to increase demand-side participation in real-time markets. Examples include expansion of real-time pricing and interruptible load programs; coincident peak pricing programs in which consumers are guaranteed to face high real-time prices only a limited number of hours per year (Uhr, 2001); and capacity subscriptions (in which consumers buy a “fuse” that limits them to certain demand levels during times of capacity shortages; Doorman, 2001).

Other proposals instead address the supply of capacity, the focus of this report. These proposals are generally viewed as interim measures to be implemented only until the demand-side is “fixed.” Some supply-side proposals involve direct payments for installed capacity, others establish capacity requirements and tradable credit systems, and still others rely on short-term operating reserves auctions to provide extra incentive for plant construction. Each proposal has its proponents who are quick to criticize other approaches. For example, tradable capacity credit systems (such as the ICAP system) are viewed by some as an unnecessary "relict" of regulation that only serves to distort energy prices and provide undeserved revenues to producers with

excess capacity. On the other hand, systems that would rely only on energy and operating reserves prices have been criticized as jeopardizing the reliability of the power system by relying on volatile markets that are distorted by strategic manipulation and price caps, and which are alleged to discourage risk-averse investors. In this report, we review the concerns that have been raised about the ability of energy markets alone to elicit optimal capacity additions. We discuss the basic principles involved in both price-based and quantity-based approaches to enhancing incentives for capacity additions, and the advantages and disadvantages that have been ascribed to each. Particular reference is made to the PJM ICAP system. We note that what are presented as polar opposites (price-based vs. quantity-based; installed capacity vs. operating reserves) actually show some convergence in the real world. The PJM ICAP system, which is nominally a quantity-based, installed capacity system, increasingly shows characteristics of the other systems. For instance, load-serving entities who have insufficient capacity credits pay a penalty, which means that at such times the system is in fact price-based instead of quantity-based. If the penalty is not too high, LSEs may choose to pay it rather than buy ICAP, and the total installed capacity in the market may fall short of the amount desired. Further, because requirements are imposed daily and because PJM is considering measures to reward capacity availability during peak times, the system may be acquiring some of the characteristics of an operating reserves-based system.

We also present a stylized model of capacity decision making for a PJM-sized market. This model simulates how competitive providers of peak and baseload capacity react in market systems to alternative approaches to dealing with the capacity question: energy spot markets alone, fixed capacity payments, ICAP-type markets, and reliance on operating reserves markets. This analysis considers how markets for energy, ICAP and/or generation reserves interact to provide incentives for capacity additions.

We reach the following conclusions about the basic approaches.

- In theory, any of the approaches can result in the provision of the socially optimal level of capacity adequacy (which we measure by the “loss of load probability”), if demand is insensitive to real-time prices. Consumers pay the same amount in each of the systems; the creation of a separate capacity market does not raise consumer costs in the long run.

- If alternate market designs give the same total capacity, the same mix of peaking and baseload capacity results from each.

- Caps upon prices of energy and operating reserves can prevent attainment of socially optimal levels of capacity in the absence of separate payments or markets for installed capacity.

- Demand responsiveness to real-time prices would imply that systems that provide price signals reflecting actual capacity scarcity (such as operating reserves markets) would be preferred to systems that average out costs over time (such as ICAP markets). Similarly, such price signals are desirable for motivating generators to increase their availability during peak periods. However, as proposed in Hobbs *et al.* (2001b), it is possible to design real-time pricing programs to send high price signals to consumers even if an ICAP system is in place and generators are subject to price caps.

- If there is a social goal to promote price stability and avoid price spikes, then the ICAP system together with price caps on energy has the advantage of lower energy price volatility. However, financial instruments (such as options or contracts for differences) can also be used to insulate power consumers from the risks associated with price fluctuations.

The general implication of these results for PJM is the following. Under certain assumptions, the analysis shows that the long run effect of an installed capacity market is to provide sufficient returns through capacity credit sales so that new generators can cover their capital costs. Among these assumptions are:

1. there is no market power (*i.e.*, companies cannot unilaterally or through coordinated action affect market prices)
2. the capacity constraint is binding and motivates additional plant construction, and
3. generators add capacity if expected revenues exceed expected costs (risk neutrality).

In theory, such returns from credit sales to generators are desirable under three conditions:

1. price caps or lack of real-time pricing for retail service keep energy prices below the consumers' value of unserved energy (VUE) during times of shortage;
2. the amount of capacity required by the ICAP rule is close to the socially optimal level (at which the marginal social benefit of additional capacity equals the marginal social cost); and
3. capacity contracts are enforceable so that the capacity is indeed available when system shortages otherwise would occur.

Under these conditions, the analysis indicates that capacity markets are just as efficient and no more costly to consumers than other ways of addressing ample capacity (fixed capacity payments, price spikes that reflect VUE, and operating reserves markets with appropriate payments).

These results imply the following for the PJM ICAP system.

- *If* the market is competitive, revenues from capacity sales replace revenues that would have been received from price spikes had energy prices been able to rise to the full VUE during shortages. Significant revenues from capacity sales are not necessarily an indication of market power, nor are they necessarily "excess." The long run cost of power is approximately the same as would occur under other systems.
- The lower the energy price cap, the greater the equilibrium capacity price will be. Thus, using price caps to depress volatility in the energy or operating reserves markets will increase the return competitive generators will need to earn in the capacity market, if adequacy targets are still to be achieved.
- If a neighboring region relies on price spikes to provide adequate incentive for capacity additions (perhaps by having no cap on prices), then during times of capacity shortages the PJM energy price will generally be less than the neighboring price. This will occur either when the PJM price hits the price cap, or when transmission capacity between the two systems is constraining. In the former case, there is a natural incentive for a generator to divert capacity from PJM to where it will earn a higher return, if it does not forfeit all the income it receives from the PJM capacity market. In so doing, the generator could get the best of both worlds: capacity payments from the PJM market for

the period when capacity is not short, and payments in excess of marginal costs from the other market when capacity is short—essentially being paid twice for the same capacity. PJM customers will then pay for the capacity, but the actual benefits will flow to other systems.

- There is evidence from the PJM Market Monitoring Unit (PJM-Interconnection, 2000d) that this diversion has been occurring to some extent in the PJM system. Several possible fixes are suggested by our analysis. One is to abandon the ICAP system entirely and either scrap the energy price cap or implement an operating reserves market whose prices would motivate capacity to stay within PJM during periods of shortage. However, if energy or reserves prices do not rise high enough or often enough to encourage sufficient capacity construction, then PJM adequacy may suffer. Another proposed fix is to replace the ICAP requirement with a requirement that load serving entities acquire sufficient call options to cover their peak demands (Oren, 2000; Vázquez *et al.*, 2001). A third possible fix is to make the capacity contracts enforceable so that if PJM load-serving entities pay for capacity during slack periods it is also available during periods of shortage. This could involve switching from a daily ICAP requirement to one that is seasonal or annual. The penalty for withdrawing capacity to sell power elsewhere could then be defined as equal to the full liquidated damages to PJM, consisting of either:

  - the cost of replacement energy supplies for the periods of time capacity is diverted, or

  - the cost of capacity credits for the entire season.

  The former cost would erase the extra gross margin that capacity would obtain from the external market, while the latter would erase the extra gross margin that would be gained from PJM. Either penalty in the long run prevents capacity from being paid twice, once by PJM and once again by the neighboring system. Either would also leave PJM whole, in theory. In the case of the first penalty, PJM would be able to buy replacement energy elsewhere. The second penalty should, in the long run, provide sufficient funds for PJM to buy capacity from other sources.

Future research should extend this analysis to analyze several issues. One is: how much demand elasticity is enough for the pure energy spot price system to induce the right level of capacity investment (Hirst and Kirby, 2000)? Our analysis above has not considered what will happen as real-time pricing and other programs increase the responsiveness of demand.

Another important issue is the implication of market power for the above conclusions. The quantitative analysis has been performed assuming perfect competition, and the results may be dramatically different if strategic generators can significantly affect energy and capacity prices. For instance, pure competition models predict that ICAP prices should be close to zero when there is excess capacity, but this has not always been the case for the PJM market. Alternative explanations for significant prices for ICAP when there is more than enough capacity include the O&M costs of maintaining availability, opportunity cost pricing based on sales opportunities in other regions, uncertainty concerning future capacity adequacy, and the exercise of market power. Both theoretical and empirical research could shed light on which of these explanations is most applicable.

Other issues that should be addressed in a quantitative way include interactions of two or more regions with different capacity market designs; the effect of alternative capacity market designs on retirement of existing units and the resulting economic and environmental implications, and whether locational capacity markets are desirable. The latter issue is important if locational marginal pricing (LMP) causes energy price differentials to persist within PJM; it is not clear whether the ICAP market together with LMP encourages capacity to be sited in the correct places. A final issue that deserves analysis is the effect of construction lead times and price expectations on investment under alternative systems.

ACRONYMS[1]

ALM	Active Load Management
CBM	Capacity Benefit Margin
CAISO	California Independent System Operator
FERC	Federal Energy Regulatory Commission
GADS	Generating Availability Data System
ICAP	Installed Capacity
ISO	Independent System Operator
ISO-NE	New-England Independent System Operator
LA	Load Aggregator
LDC	Local Distribution Company
LOLP	Loss of Load Probability
LMP	Locational Marginal Pricing
LSE	Load Serving Entity
NEPOOL	New England Energy Pool
NERC	North American Electric Reliability Council
NYISO	New York Independent System Operator
OA	Operating Agreement
PJM	Pennsylvania, New Jersey and Maryland Power Pool
PJM-ISO	Independent System Operator
PJM-OI	PJM's Office of the Interconnection
RTO	Regional Transmission Organization
RAA	Reliability Assurance Agreement
VUE	Value of Unserved Energy (also called value of lost load, VOLL)
WTP	Consumer Willingness to Pay

[1] Definitions of technical terms can be found in Section 1.1 and Appendix I.

1. INTRODUCTION

The goal of this report is to present an overview of issues concerning the market for capacity in deregulated power markets with a focus on the PJM ICAP market.

The PJM Independent System Operator (ISO) has administered the PJM Open Access Transmission Tariff since April 1, 1997 and the PJM Energy Market since January 1, 1998 in Delaware, Maryland, New Jersey, Pennsylvania, Virginia and the District of Columbia. The ISO clears daily and hourly energy markets; clears daily and monthly installed capacity credit (ICAP) markets; and buys necessary ancillary services. The capacity requirements evolved from operating rules of the former PJM power pool, established in 1927. PJM serves 8% of the US population with a pooled generating capacity of 56,000 MW, generating 2.6×10^8 MWh in 1998, 8.5% of the total electric energy produced in US in that year (EIA/DOE, 1999). The PJM Office of the Interconnection (PJM-OI) compiles a load forecast for the pool from forecasts by individual local distribution companies (LDCs) for the year's Planning Period, and also establishes the PJM Reserve Requirement. Based upon this requirement, PJM OI then allocates to every load serving entity (LSE) in the Pool its generating capacity obligations as a share of the total PJM Control Area generating capacity requirement. Appendix II describes these PJM rules in more detail.

The design of the PJM ICAP market has been controversial for two basic reasons. One has been perceived inadequacies in its operation: some LSEs declined to sell their extra capacity in the market in 1999, while since the late spring of 2000 there have very high and volatile prices for credits. The second reason is more fundamental: there is disagreement over what capacity market mechanisms—if any—are needed for ensuring generation adequacy (Rau, 1999). Some power markets, such as California's, have no installed capacity market at all, while other markets use payments or other mechanisms to provide incentives for installing capacity. The following general questions represent the key issues about capacity markets that are addressed in this report:

1. Under what conditions, if any, can energy markets alone ensure adequate investment in generation capacity?
2. Under what conditions can alternative mechanisms (such as capacity (ICAP) markets, payments for capacity, and markets for operating reserves) ensure adequate generation investment?
3. What are the advantages and disadvantages of different systems for providing incentives for generation capacity investment?

This report summarizes various perspectives concerning the issue of how and whether markets for capacity should be implemented. In this introductory section, we first provide some basic definitions of terms used in the report, and then give a general overview of the debate. The history of the PJM ICAP system is then summarized, focusing on its roots in the capacity reserve requirements of the old PJM pool agreement. Finally, §1.4 presents an overview of the rest of the report.

1.1. Definitions

Below are definitions of some key terms that are frequently used in this report. Additional terms are defined in Appendix I.

Reliability: According to NERC (Hirst *et al.*, 1999), reliability is defined as "the degree to which the performance of the elements of the system results in power being delivered to consumers within accepted standards and the amount desired." It refers to two concepts: *adequacy* and *security*, defined as follows (*ibid*.).

> *Adequacy*: "the ability of the system to supply the aggregate electric power and energy requirements of the consumer at all times."
>
> *Security*: "the ability of the system to withstand sudden disturbances."

Adequacy is said to deal with future planning and investment so there is sufficient generation and transmission resources to meet projected needs (plus reserves). Security deals with the short run operations, ensuring that the system will remain intact after equipment outages, changes in demands, or other disturbances occur. The focus of installed capacity markets is upon long-run *adequacy* of capacity resources.

Load Serving Entity or **LSE** means any entity (or the duly designated agent of such an entity), including a load aggregator or power marketer that (i) serves end-users within the PJM Control Area and (ii) has been granted the authority or has an obligation pursuant to state or local law, regulation or franchise to sell electric energy to end-users located within the PJM Control Area.

Capacity generally refers to a product or service that contributes to adequacy even when it is not needed to satisfy instantaneous demand (Singh and Jacobs, 2000).

Long term reserves, often called planning reserves, are available power sources considered in the long-term projections performance of the system. They provide long-term insurance against problems that might otherwise arise when units are not available, and allow for unanticipated long-term load growth (Hirst *et al.*, 1999).

Short term reserves, also called operating reserves, are energy sources available and fully operational in the system within a short period of time (10, 30 or 60 minutes) from the moment they are ordered to dispatch. Operating reserves consist of two types (Hirst and Hadley, 1999): spinning reserves (on-line capacity synchronized to the grid, capable of increasing full output within 10 minutes), and supplemental reserves (which do not need to be synchronized to the grid and can be available within 30 - 60 minutes). NERC requires that utilities maintain a minimum of 4-8% of the projected daily peak as operating reserves (Hirst *et al.*, 1999).

Unforced Capacity is the installed capacity rated at summer conditions, derated to account for the likelihood of experiencing a forced outage or forced derating. PJM calculates the derating adjustment for each capacity resource on a rolling 12-month average (which shall be updated each month for the 12 months ending two months prior to the billing month).

PJM Reserve Margin: the percentage of the forecast pool resource requirement for the planning period over and above the coincident forecast peak of all LSEs. The PJM planning period is defined as the twelve months beginning June 1 and extending through May 31st of the following year.

IRM – ICAP Requirement: the level of installed reserve capacity needed to maintain the desired level of adequacy. In PJM, IRM is determined as determined by the PJM OI, and is based on the desired reserve requirement.

Ancillary Services: Six distinct services identified by the Federal Energy Regulatory Commission (FERC) Order 888 as necessary for providing electricity service to consumers:

1. Scheduling, system control and dispatch service;
2. Reactive supply and voltage control from generation source service;
3. Regulation and frequency response service;
4. Energy imbalance service;
5. Operating reserve-spinning reserve service (system protection); and
6. Operating reserve-supplemental reserve service.

Loss compensation services may also be provided by a transmission provider as an ancillary service to transmission customers. Ancillary costs have been estimated as representing from 5 to 25% of the bulk power costs of utilities in 1996 (Earle *et al.*, 1999).

1.2. Problem Statement

For most of the 20th century, the generation, transmission, distribution, and control of electricity supply were regulated by state and federal law under the presumption of being a natural monopoly. It is generally assumed that transmission and distribution remain natural monopolies. However, this is no longer true for generation, due to increased market sizes and weakened scale economies resulting from technological innovation. Cost-plus price regulation of generation, along with elaborate licensing procedures for new power plants have come to be viewed as unnecessary and discouraging of innovation and efficiency. Thus, restructuring of the generation sector has proceeded with the goals of achieving short-term (pricing and trading) and long-term (investment) economic efficiency as an outcome of free decisions taken by consumers and producers.

Traditional utility regulation allowed the electric industry to recover the full costs of investments required to achieve adequacy goals. The widely accepted reliability goal of “1 day in 10 years” could be attained by requiring utilities to add capacity whose cost would then be added to the rate-base and recovered through prices set by the utility commission. In contrast, under restructuring, price is not fixed by an average costing model, but follows more closely the marginal cost of providing the service. But as a result, generators cannot count on being able to recover the full cost of installing generation capacity. As a result, there are concerns about whether there will be adequate capacity. For instance, Hirst and Hadley (1999) report that reserve margins across the US dropped from 22% in 1990 to 16% in 1998. These decreasing margins are in spite of significant planned and actual capacity additions by independent power producers.

On the other hand, the North American Electric Reliability Council (NERC) now projects that capacity margins will be adequate for the mid-Atlantic region (Table 1.1). Projections they made in previous years had resulted in worries about whether there will be sufficient adequacy. For instance, they had earlier projected a reserve margin of 5.1% for 2007 (NERC, 1998).

Table 1.1: NERC Projections for the Mid-Atlantic Reliability Council (MAAC)

	Summer		Winter	
	2000	2004	2000/1	2004/5
Net Internal Demand [MW]	49,325	52,406	42,307	44,933
Net Capacity [MW]	57,831	74,496	60,815	79,309
Installed Reserve Margin [%]	14.7	29.7	30.4	43.3

Source: NERC, Reliability Assessment 2000-2009, October 2000

Throughout the world, different strategies have been designed to cope with the problem of ensuring generation adequacy. In PJM, the approach has been to require LSEs to either own installed capacity or acquire installed capacity (ICAP) credits to meet an ICAP requirement set by the PJM ISO. (See Appendix II for a detailed explanation of the PJM system.) As a result, separate markets for capacity have been created in parallel to the energy market. ICAP credits can be traded bilaterally, or can be sold through the PJM ICAP monthly and daily markets; as a result, capacity is valued, and incentives are provided for adding capacity in a timely fashion.

But will the ICAP system provide the socially optimum level of reliability? Is it even necessary? What are the advantages of other means of providing incentives for investing in generation capacity? These issues have been of concern to industry and researchers in the past two decades. The theory of electricity spot pricing (Schweppe *et al.*, 1982, 1987; Caramanis, 1982) states that correctly calculated spot prices for energy should be sufficient to motivate the optimal (social net benefits maximizing) amount of capacity additions; thus, a separate capacity credit market should not be necessary in such markets.

However, this theory makes certain restrictive assumptions. The key assumption is that the market clears at a price that equates the consumers' WTP for power (or to avoid loss of load) with the marginal cost of supply. When capacity is insufficient, prices should rise high enough in real time so that consumers are motivated to cut back their quantity demanded.[2] Jaffe and Felder (1996) point out that the ongoing debate about whether or not to have a separate capacity market arises largely because the market for electricity, unlike markets for other commodities, does not actually clear in this manner.

However, even though the energy market fails to clear according to theory, some observers argue that an installed capacity market is still unnecessary. Singh and Jacobs (2000), Cramton and Lien (2000), Graves and Read (1998) and others argue that the notion of a capacity product distinct from energy is a concept that is left over from the regulatory era, and is not needed to create the appropriate signals for the electric energy system to have socially optimum adequacy. This view is shared by some USDOE staff who have expressed the view that separate capacity markets are a transition stage on the road to fully competitive markets. Singh and Jacobs (2000)

[2] To account for the fact that demand cannot instantaneously respond to price and that equipment outages can occur with no warning, the theory of spot pricing calculates prices in a more sophisticated manner. When a significant risk of capacity shortages arises, spot prices increase to account for the probability that a loss of load will occur and the resulting loss of consumer benefits (Schweppe *et al.*, 1987).

suggest that operating reserve markets will instead provide sufficient incentive for capacity additions.

Because of the rising concern over system adequacy and the existence of competing proposals for dealing with it, the need for and appropriate design of capacity markets is controversial. Among the questions that are debated are:

- In the absence of a separate market for a capacity, can a deregulated energy market be expected to provide the socially preferred level of adequacy? Why or why not?
- What general incentives do alternative capacity market designs provide for additions or retirements of capacity of various types, fuels, and sizes? Can alternative designs achieve socially efficient levels of adequacy?
- What might the implications be of alternative capacity market designs for generation mix, market concentration, costs, prices, and emissions?
- Might alternative capacity market designs affect the ability of restructured power markets to achieve the goals of low prices and technological innovation (Shepherd, 1998)?
- Can different designs for capacity markets coexist in adjacent regions? (Stoft (2000a) argues that the PJM ICAP market design will result in diversion of capacity during peak periods to neighboring regions lacking such markets.)

This report is a review of the debate over these issues. Our approach is to summarize the literature, and then to use a simple model to illustrate the theoretical points about the relative efficiency of different capacity market designs.

1.3. The Capacity Market in PJM

Between 1974 and 1998, capacity adequacy in the PJM Power Pool was regulated by the old PJM Interconnection Agreement (PJM Installed Capacity Accounting, Schedules 2, 3 and 4) and the PJM Reserve Sharing Agreement (RSA). The goal was to govern installed capacity reserve sharing obligations in the pool. The ability of PJM members to pool their resources for purposes of reserve sharing had generated significant reliability benefits and cost savings for the PJM members over the years. Only LSEs (then called “utilities”) had the obligation to comply with this agreement and the allocation of capacity required of each LSE. The intent of this rule was to require that sufficient capacity would be built, and to prevent a utility from “leaning on the pool” by shirking its responsibility to contribute capacity resources and thereby imposing a reliability problem on all of PJM (FERC, 1997).

After restructuring of PJM in 1997, the capacity reserve sharing mechanism continued to be regulated by the previous scheme. During the transition period towards retail choice, the reserve calculations were performed in two steps: (1) the Forecast LSE Obligation was computed for all of PJM; and (2) each Forecast LSE Obligation was then allocated among the LSEs within PJM. Each PJM member was obliged to demonstrate that it had sufficient installed capacity (ICAP) to meet its forecast LSE obligation over the following two years. Under the Reliability Agreement, each LSE had to submit a plan indicating how it intended to meet its Forecast LSE Obligation over the following 24 months for loads that it served under its franchise. In any billing month, if an LSE failed to meet its installed capacity obligation, it was assessed the capacity deficiency charge.

Then in 1998, the PJM ISO organized a working group, the Visible Capacity Market Users Group, (PJM-Interconnection, 1998d) and asked it to modify the pool rules to update the previous ICAP system to accommodate the introduction of retail choice in PJM. The group met between May and October 1998. Its recommended modifications were approved by the Reliability Committee, and filed with FERC on October 14th, 1998 (PJM-Interconnection 1998a). The reform was introduced as a modification of the Schedule 11 in the PJM ISO Operating Agreement under which PJM was deregulated.

In order to implement a reserve sharing approach that would reduce the cost of installed capacity reserves, PJM developed a set of specific procedures for: (1) determining the pool-wide generation requirement needed to meet pool-wide loads, including reserves; (2) determining each member's individual obligation to contribute to the pool-wide generation requirement; (3) measuring each member's compliance with its obligation; and (4) developing charges that apply whenever a member fails to meet its individual obligation (referred to as a capacity deficiency). The modification to Schedule 11 also allowed non-LSEs to participate as parties to the Reliability Agreement.

Several characteristics of the current ICAP market were carried over from the previous scheme. First, the capacity deficiency charge was billed on a daily basis, based to an annual charge of $58.40/kW/year. The capacity deficiency charge is based on the cost of installing a combustion turbine. Second, the procedures used to set obligations (installed reserve margins, etc.) were annual processes based on a planning year (June-May) concept. The two-year period obligation to meet those parameters was eliminated when the ICAP market was implemented in 1999. Third, the semi-annual unit capability testing and reporting requirements, which are imposed on the owners of generating facilities, is also a procedure inherited from the previous rules and is designed to keep track of the state of the system. Fourth, the Capacity Benefit Margin, a transmission import capability reserved by the ISO to support the system in case of reduced installed reserve margins, was also carried over to the new design. Finally, the revenues from any capacity deficiency charges are distributed to LSEs that maintain installed capacity in excess of their Forecast LSE Obligation, and other owners of Capacity Resources with excess capacity as in the previous design (PJM-Interconnection, 1997e).

An important motivation for establishing the ICAP market was to assure capacity owners that they would recover their fixed costs. During discussions of the ICAP market design in the PJM Public Interest & Environmental Organization User Group May 7, 1999 meeting (PJM-Interconnection, 1999a), a participant questioned whether the use of Operating Reserves, an approach used in ECAR, would do a better job than ICAP in maintaining adequacy. In reply, a staff member from PJM said "to have operating reserves on a daily basis it is necessary to have installed capacity on an annual basis." Implementing an installed capacity approach would also ensure the recovery of fixed costs as well as operating costs, he noted. In response to the suggestion that fixed costs should be recovered in energy costs where market based rates are used, a staff member suggested that "energy costs must be demonstrated to exceed marginal costs, that history is important, and that it takes time to develop consensus and then implement the changes." These responses indicate there existed a motivation for the ICAP market to avoid the disruption that could accompany too drastic a departure from previous practices, and that it was important to ensure fixed cost recovery in the turbulent initial years in which the market would operate.

PJM started the Future Adequacy Working Group (FAWG) in 1999 to address short-term modifications to the adequacy model under which PJM operates. Since August 2000, the group has been considering several proposals to modify or replace the current ICAP model. The group seeks to correct distortions in the energy market that several stakeholders claim are introduced by ICAP (PJM-Interconnection, 2000e).

PJM is also expanding to encompass a second control area by January 2002, the so-called "PJM West." This is expected to consist of Duquense Light Co. and Allegheny Power. In a filing to FERC March 14th, 2001 (PJM Interconnection, 2001a), PJM proposed a structure for PJM-West consisting of the PJM energy market extension to the western region, along with the LMP and the congestion management systems. The capacity requirements for LSEs are going to be in the form of Available Capacity Resources set initially to 106% of their day ahead peak load. This proposal is a mixture of a pure spinning/operating reserve and traditional ICAP requirements. If accepted, this proposal means that PJM will have a dual system: a pure ICAP for the PJM "East" area and a mixed system for PJM-West.

1.4. Scope of this Report

In §2, we begin with a brief review of adequacy concepts, and the basic theory of spot pricing. The assumptions of the ideal spot pricing market are reviewed, and the implications of their violation discussed. In §3, some alternative capacity markets are defined, including installed capacity markets, installed capacity payments, and operating reserves markets. Their ability to overcome the market failures that may prevent spot markets from achieving sufficient adequacy is discussed. General advantages and disadvantages of these alternative designs are reviewed. In §4, results from a simple reliability market model are used to illustrate the operation of the different market designs. The model results indicate that each of the alternative designs can economically achieve desired adequacy levels under certain assumptions, and that ICAP does not necessarily distort the market or lead to excess revenues for generators. Finally, §5 summarizes our answers to the following six questions:

1. Is ICAP a fictional product with no inherent value?
2. Does ICAP distort energy prices?
3. Is ICAP insufficiently related to system reliability compared to operating reserves?
4. Is ICAP subject to migration, and so cannot ensure adequacy?
5. Does ICAP distort entry/exit decisions?
6. Does ICAP magnify market power problems?

Several appendices are included for reference. Appendix I presents additional definitions of terms used to describe energy and capacity markets. Appendices II-IV provide more detailed information on reliability calculations and market operations in the PJM ICAP system. Appendix II reviews the PJM reserves planning process that determines the ICAP requirements that each load serving entity must meet. Appendix III describes the PJM capacity deficiency and recall process, which is important in determining whether other regions can "poach" PJM capacity at critical times. Appendix IV summarizes data concerning operating parameters relevant to the PJM ICAP market and historical capacity price and transaction data. Finally, Appendix V gives a summary of possible ICAP market gaming reported by the PJM Market Monitoring Unit (PJM-Interconnection 2000d).

2. CAPACITY RELIABILITY AND ELECTRICITY SPOT MARKETS

2.1. Electric Power System Adequacy Definitions and Calculations

Four basic indices are used to describe the adequacy of generation system capacity (Billinton and Allan, 1992):

- Probability of capacity shortage P. This is the expected fraction of time (or hours/year) that system generation capacity is less than total demand. By "expected", we mean the probability-weighted average. This is what PJM refers to as its "loss of load probability" (LOLP), and is also known as a loss of load expectation (LOLE). PJM's target of 1 day in 10 years literally means a probability of 0.000274 = (1 day)/(365 days * 10 years) (R. Gramlich, formerly at PJM, personal communication).
- Expected unserved energy EUE. This is the expected MWh per period of time (*e.g.*, MWh/year) of demand that is unmet because of capacity shortages.
- Expected frequency of outage F. This is the expected frequency of occurrence of outage events (*e.g.*, per year). Some users of the term LOLP are actually referring to a frequency (*e.g.*, "one outage occurring every 10 years, on average").
- Expected duration of outage D. This is the average length of a period of capacity shortage, usually expressed in hours.

Three of the indices are related by the equation P [hours/year] = D [hours] * F [1/year]. EUE [MWh/year] in turn equal P [hours/year]*LOL [MW], where LOL is the average MW of load lost during capacity shortages. P and EUE are generally calculated by either "convolution" or Monte Carlo methods that consider both the distribution of demand and the probability ("forced outage rate") of generator unplanned outages (Billinton and Allan, 1992). More sophisticated methods consider the effect of transmission bottlenecks and planned maintenance outages upon P and EUE. PJM's reliability modeling is of this type (PJM-Interconnection, 1998c). Frequency and duration statistics are less commonly calculated, and are obtained either by Monte Carlo methods or so-called "frequency-duration" techniques based upon Markov processes. The GUS Manual (PJM-Interconnection, 2001j) uses frequency-duration methods to model the states of individual generators, but not the entire system.

Power system planners and regulators use these statistics and methods to assess the reliability of existing power systems; to estimate the effect of resource additions upon reliability; and to define capacity reserve margins necessary to obtain a target reliability level. Thus, for example, PJM calculates the installed capacity necessary to ensure that the system peak during a given planning period will be met with an assumed loss of load probability.

Later in this issues paper, we use simple convolution-based methods to assess the P associated with capacity mixes that could result from different capacity market designs. The system analyzed is an idealized PJM-like system in which there are no transmission constraints and two types of generating plants (base load and peaking).

2.2. Spot Markets for Energy and Capacity Decisions

2.2.1. Theoretical Ideal

In the absence of “market failure,” market forces that equate marginal private costs with marginal private benefits will also equate marginal social benefits and marginal social costs. Microeconomic theory shows that such markets will allocate resources efficiently. By “efficiently”, what economists mean is that it would not be possible to reallocate resources without making at least one person in society worse off.

This ideal model of market function—together with political pressure from large customers hoping for lower rates—has inspired the deregulation of the electric markets throughout the world. The main goals of restructuring can be stated as follows (Wolak, 1998):

- Economic efficiency, including allocative efficiency (price = marginal benefit to consumers = marginal cost to producers), production efficiency (outputs produced at least social cost), and dynamic efficiency (technological progress). These translate into both short-term efficiency of pricing, trade, and dispatch, and optimal long-term investments.

- Equity: assignment of costs to those responsible for them through the marginal cost-based prices that arise from competitive markets

- Appropriate consideration of externalities and public goods

Spot pricing theory provides a consistent and idealized description of how an unregulated power supply system could result in efficient operation, investment and pricing. The fundamental concept was first suggested in 1971 by the Nobel Prize winning economist William Vickrey (1971) in a paper called “Responsive Pricing of Utility Services.” The concept was later applied to power systems by Fred Schweppe (1978) in a paper entitled “Power Systems 2000" and more fully elaborated in later papers and a book (Schweppe *et al.*, 1987). Caramanis (1982) describes specifically how the spot pricing ideas apply to capacity investment.[3]

Spot pricing theory describes an energy market in which price is free to rise and fall in near-real time in order to clear the market, and customers and generators both see and respond to prices. Such prices are also known as real-time or responsive prices. It turns out that, in theory, such a market would optimize “social net benefits,” defined as the difference between the value of electricity usage and the cost of supply. No external reliability condition is imposed on the system; instead reliability is optimized by interplay of the consumers' willingness to pay for energy at different times and generators' willingness to provide it.

Under this approach, the spot price for buying and selling electric energy is determined by the supply and demand conditions at each instant.[4] This results in a pricing system that does not require the complications of block rates, demand charges, back-up charges, or capacity credits.

---

[3] Green (2000) gives a summary of the theory and its implications for the electricity market.

[4] See Note 2 above. A general statement of the theory also covers the situation in which there are two groups of consumers, one subject to spot prices and a second that faces predetermined prices. The latter group may have to be rationed under certain events; the theory assumes (naively) that the system operator knows their outage costs and cuts service to customers for whom the outage costs are smallest.

More general versions of spot pricing theory have been developed, include the locational marginal pricing system used by PJM (Schweppe *et al.*, 1987) and reliability-differentiated pricing (Chao and Wilson, 1987).

Focusing on long-term efficiency, Caramanis (1982) developed a model of generation capacity decisions under spot pricing. He shows that profit maximizing behavior results in the level of investment in capacity that also maximizes social net benefits. Generating facilities earn enough during periods of high prices to just cover their capital costs. For peaking facilities, prices exceed their marginal operating costs during times of capacity shortage when the price increases in order to ration supply to the most valuable uses. At such times, the price equals the marginal willingness to pay for power by consumers who are going without energy, which may be well above the marginal cost of supply.

The basic assumptions underlying these important results are the following: [5]

1. Real-time prices:
   - Both producers and consumers see the real-time market price, and make production and consumption decisions, respectively, in order to maximize their net benefits. When capacity shortages occur, price rises, and customers whose WTP for power (equivalent to the value of unserved energy, VUE) is less than price decrease their consumption, while those customers who have a higher value for power pay the price and continue to consume. As a result, the spot price will equal the VUE for the marginal use that is cut off in times of shortage.[6]
   - There are no government imposed taxes, charges, or restrictions on price movements
2. No participant (consumer or producer) has market power. As a result, consumers adjust his or her consumption until their VUE equals the market price. Likewise, each producer equates his or her marginal cost of production with the market price.
3. There is perfect knowledge of future prices or their distribution.
4. There are no environmental or other externalities that would cause social costs of power provision to deviate significantly from private costs.

As a result of these assumptions, as Schweppe *et al.* (1982) describes it, "the investment criterion for social maximization is the same as for a private, profit maximizing firm. For example, customers will invest in self-generation if it lowers their total cost. Under optimal spot pricing it will lower their total cost if and only if it lowers total costs for the utility as a whole." Similarly, generation capacity expansion will take place until the point where the last increment of capacity operates just enough hours per year so that its revenue from price spikes (when price equals

---

[5] When it is argued that capacity markets are not necessary because energy price spikes will provide sufficient revenue for capacity expansion (*e.g.*, Hanger *et al.*, 2000), it is being assumed that these assumptions are satisfied and that the market failures described in the next section are not significant.

[6] In the case where the ISO instead decides what customers are to be cut off (see the footnote on the previous page), the ISO would pay the marginal VUE to the market. This VUE should reflect the consequences of shortages at a given time, and the consumers' willingness to pay to avoid those consequences. The issue of how VUE should be calculated administratively is a vexed one, and there is a tremendous literature on the topic (*e.g.,* Billinton, 1994). Questionnaires involving hypothetical outages are frequently used. Some argue (*e.g.,* Chao and Peck, 1998) that demand-side bidding is the most reliable way to infer VUE as it would be based on actual choices by consumers.

VUE) just covers its capacity and operating costs. The benefit of any capacity over and above that level, in terms of avoided customer shortage costs, would be less than the expense of providing that capacity. As a result, net social benefits are maximized.

The idealized market therefore dynamically adjusts capacity. As Jaffe and Felder (1996) put it, investors will likely make their investments on the basis of expected present value of future spot prices. In general, if prices are above their optimal levels because there is too little capacity, capacity additions will be profitable and will occur. On the other hand, if there is too much installed capacity, price spikes occur too infrequently, and this encourages capacity retirements and discourages replacement.

However, if any of the above four assumptions are significantly violated, a "market failure" results, and the socially optimal amount of capacity adequacy may not be attained. The market failures can cause prices to be distorted, so that they do not represent consumers' marginal willingness to pay for reliability (VUE). Yet even in the presence of appropriate price signals, other types of failures can also discourage producers from making appropriate investments. In the next section, we summarize several important types of market failures that prevent the pure spot market envisioned by Schweppe *et al.* (1987) from providing the correct amount of generation capacity.

2.2.2. Market Failures

A market failure is usually defined as an inefficiency that arises because one or more of the assumptions of the model of perfect competition are violated. In general, the reasons for failure of the spot market for power can be grouped into four general classes of market failure described by microeconomic theory (*e.g.*, Kahn, 1998):

- imperfect information;
- externalities and public goods;
- imperfect competition (market power); and
- price rigidities resulting from government intervention and market design.

If any of these failures occur, it cannot be expected that market forces of supply and demand will equate the marginal social benefit of capacity with its marginal social cost. Each category of failure is explained below. We pay particular attention to features of power markets that lead to these market failures. In §3, we describe how alternative market designs attempt to address and correct these failures.

2.2.2.1 Imperfect information. There are two types of information imperfections that we consider. The first is the incompleteness of electricity markets: prices do not reflect consumer WTP for reliability (*i.e.*, VUE) because consumers do not see and cannot react to real-time prices. As a result, prices do not convey correct information about the social value of capacity to producers. A second information failure arises from uncertainty of future energy prices, including uncertainties arising from changes in policy and market design.

The first type of information imperfection arises when some group of market participants (consumer, producers or both) does not know the true social cost or benefits associated with a good or activity because prices are incorrect. This is the first assumption of §2.2.1. Basically, both supply and demand should be able to see and react to the real-time market price. However, because of transaction costs, the vast majority of consumers do not see a real-time price—in this sense, the market is "one sided" (Singh and Jacobs, 2000). Furthermore, even if consumers were forced to pay real-time prices, the time scale of system disturbances is such that they generally cannot react sufficiently fast to clear the market if demand unexpectedly increased or if a generator or transmission line failed. There is also, in most cases, little or no storage to buffer sudden changes in supply-demand balances. Thus, sudden price rises cannot be depended upon to induce consumers to cut back; markets will not clear; and load shedding actions will not necessarily cut off the consumers who value power the least (Jaffe and Felder, 1996). As a result, real-time prices will not rise to the VUE of the last customer whose load is shed.

Therefore, consumers do not pay the applicable VUE during energy shortages. Prices may instead be lower (or higher) than VUE. In some cases where an LSE has the ability to interrupt service, prices during times of shortage may instead be based on what a LSE's management is willing to pay to avoid having to cut off customers. (Stoft (2000a) refers to management's willingness to "stomach" rotating blackouts rather than pay very high prices during shortages.) Eventually, where LSEs in a market can choose different levels of reliability, it is conceivable that management's WTP for power may reflect that of consumers because consumers may pick their LSE based in part on its reliability.[7] However, this is manifestly not the case now, and political considerations (such as widely publicized criticisms from the mayors of New York City and Chicago during the summers of 1998 and 1999) and artificial price caps (such as PJM's $1000/MWh cap on energy bids) may be just as or even more important. Management WTP to avoid shortages may therefore overstate or understate consumer WTP for reliability.[8]

If prices systematically understate the VUE during price spikes (either because LSE management is too willing to live with shortages rather than buy expensive spot power or because of price caps), then too little investment in capacity will take place. This is the argument of Jaffe and Felder (1996) and others who argue that pure energy markets will provide insufficient incentive to expand capacity. On the other hand, prices during shortages might overstate the value of reliability. Demand will apparently be extremely inelastic, as high spot prices will not stimulate demand reductions by customers—as most customers do not see real-time prices. Political factors may push prices artificially high, and too much capacity and, thus, reliability might be provided in the long term. The possibility of prices being too high rather than too low should not be dismissed. For instance, few LSEs have taken full advantage of opportunities for direct load control; the reduction in loads such measures can accomplish could be less costly than power purchase costs of $1000/MWh and up that have been experienced during price spikes (Kueck *et al.*, 2001).

The second information-related failure concerns the information that investors need to make a decision. Even if prices appropriately reflect consumer WTP for reliability, suppliers might not respond to those signals in a socially optimal way. Price uncertainty could discourage risk

---

[7]However, in PJM and many other markets, the LSE has no ability to interrupt its customers just because prices are too high. In the case of PJM, this is because the ISO makes interruption decisions and not the LSE. In other cases, retail access means that a given area will have customers served by more than one LSE, and metering and control technologies are not installed that would allow just one LSE's customers to be interrupted.

[8] Although it is implicitly assumed in many articles that prices during shortages are less than consumer VUE and thus provide too little incentive for capacity additions, the opposite is arguably just as likely, as is argued below.

averse suppliers from investing (Jaffe and Felder, 1996; Virginia State Commission, 1997).[9] The price volatility inherent in a system without significant storage by itself can discourage investment. For instance, Loehr (1998, cited by Hirst *et al.* 1999) argues that in a free market nobody would build peakers that would be profitably used only 100 to 200 hours per year, even if prices are very high at those times. (However, such plants have been proposed and built by independent providers in PJM!) Magnifying the importance of price volatility is that the frequency of price spikes itself is highly uncertain. This because their frequency depends upon forced outage rates (which are uncertain and may change in the future under restructuring, Hirst *et al.*, 1999) and, especially, behavioral factors such as exercise of market power or changes in market design and price caps.

Another example of a behavioral uncertainty is the future of real-time or "responsive" pricing for large commercial and industrial customers. Schweppe in his various writings (*e.g.*, Schweppe, 1978) envisioned that such pricing would become more widely used over time, making demands more price responsive and decreasing the frequency of price spikes. In reality, the immediate effect of restructuring has been to discourage responsive pricing experiments by utilities (such as BGE) because of uncertainties concerning cost recovery and because customers have expressed a preference for more predictable prices. However, the metering costs of real-time pricing continue to fall, and its potential benefits are high enough that this trend might be reversed soon (Jaffe and Felder, 1996).[10] Cameron and Cramton (1999) argue that this trend will also be

---

[9] Price volatility could also encourage investment in capacity by risk averse LSEs who wish to avoid price spikes. Capacity in that case can be viewed as a "put option" that can be used to lower the upside risks of price variation. Wu *et al.* (2000) provide a theoretical framework for analyzing advance contracting vs. spot market purchase decisions.

[10]According to D. Hale (personal communication, 2000), experiments using the USDOE National Energy Modeling System show that having 10% of the load in the eastern US sensitive to real-time prices can dramatically change price patterns, reducing prices under peak load conditions. But it might not be practical to have such a large fraction of the load be responsive within a 10 minute market window, with is the goal stated by the PJM ISO CEO at the 2000 IEEE Reliability Symposium (IEEE-USA, 2000).

Hirst and Kirby (2000) have written a thorough report on the state of the art in retail-load participation in competitive electricity markets in the US. The authors ask why so little is happening, given the large benefits of having a price-responsive demand. In their report they enumerate barriers to the full implementation of dynamic pricing in the US for customers. One of the barriers is the regulatory framework that most ISOs have implemented, reflecting indecisiveness on how to create truly competitive markets while also protecting retail customers. The regulations that have been adopted have generally isolated consumers from dynamic market price signals. Hirst and Kirby (2000) argue that energy prices for consumers should reflect the risk of energy fluctuations, or the costs of managing that risk. Regulatory uncertainty has also inflated the risks of utility investments in the metering and related services required to implement real-time pricing. The uncertainty arises from the fact that it is often unclear as to who will own the metering systems and the data collected after utilities have made the investments.

The second set of barriers includes technical ones that act against the easy proliferation of these systems. First is that scale economies have not yet been achieved in the production of metering systems, so their costs have remained high. There is also a need for technical standards so the components can integrate easily to the system.

The third set of barriers encompasses institutional attitudes that limit the adoption of these practices. Hirst and Kirby (2000) enumerate a number of these attitudes. Examples include: the belief that energy prices should be time invariant; the lack of recognition of the risk management component of electricity pricing; and the supply-side focus of ISOs.

encouraged by PJM's shift to a multi-settlement system, which will increase opportunities for demand-side bidding[11].

As a consequence of such uncertainties, price spikes might occur more or less often in the future. Poor information about future frequencies of price spikes will generally discourage investment in peaking capacity by independent producers, and could decrease the reliability of the power system in the absence of other mechanisms to motivate investment.

How might market designers address the failure of market prices to communicate consumer VUE or to motivate risk averse suppliers to add capacity? If under-investment is the problem, four general fixes to this market failure have often been proposed, the last three of which are discussed in more detail in §3, below.[12] The first fix is to correct the market failure itself by increasing participation by the demand-side in the real-time market. Enhanced participation would manifest itself as an increased price-elasticity of demand on a daily, hourly, or even minute time scale (Hirst, Kirby, and Hadley, 1999). Improved communications and metering technology would support this goal, as would greater incentives for customers to sign up with real-time pricing programs. LSE-sponsored load management programs (such as air conditioner and water heater load controllers) would also contribute, and are already given credit by the PJM ICAP system. A measure that would encourage LSEs to adopt such programs would a requirement that LSEs with a balance deficit during periods of shortage make compensatory payments to consumers that are subjected to involuntary outages (Soder, 2000). These payments could be based on estimates of the value of unserved demand for different customer classes.

The other fixes are instead directed at creating a market for capacity to augment the energy market. The second fix would be to impose a requirement upon LSEs to construct or contract for sufficient capacity, and to allow capacity credits to be traded. If "sufficient" were to be defined as the socially optimal amount of capacity (based of a balancing of estimated outage events and capacity costs), then correct price signals could in theory be given by the capacity credit market (see §4, below). A third fix would be provide capacity payments that reflect VUE in order to increase returns received by investors. Yet a fourth fix, argued for by Singh and Jacobs (2000), is to pay for operating reserves; price spikes that occur when reserves are in short supply could then provide the needed incentive for long-term investment.

2.2.2.2. Externalities and Public Goods. Externalities refer to spillover costs or benefits—unintended side effects of market transactions experienced by third parties. Baumol and Oates provide the following more formal definition (quoted in Kahn, 1998):

> "An externality is present whenever some individual's utility or production relationship include real (that is non-monetary) variables, whose values are chosen by others (persons, corporations, governments) without particular attention to the effects on that individual's welfare."

Public goods are goods or services that present non-excludability and non-rivalry characteristics. Non-excludability refers to the inability to exclude any consumer or generator from using it, if

---

[11] The Members Committee (PJM Interconnection, 2001b) is reviewing the final proposal for the future implementation of a PJM 2001-2002 Load-Response Pilot Program, which addresses demand-side bidding.

[12] Note that if price spikes are too *high* during shortages rather than too low—resulting in too much investment rather than too little—then none of these market changes would correct that market failure.

one of them has the possibility to consume it. Non-rivalry refers to the fact that an individual's consumption of the good or service does not diminish the amount of that good available for others. Either the presence of externalities or the existence of a public good or service result in the violation of the fourth assumption of the ideal spot market (§2.2.1).

An alternative way to view the problem of inappropriate prices described in §2.2.2.1, above, is as an externality or public good problem (Jaffe and Felder, 1996; Oren, 2000; R.P. O'Neill, Office of Economic Advisor, FERC, personal communication, 2000). Because prices may understate consumer VUE during shortages and because consumers do not pay the real-time price, the addition of capacity in order to increase system reliability can be viewed as being a public good. The reason is that all consumers will benefit from a capacity addition without necessarily paying for it. If energy prices are too low because of a price cap or other reasons, owners do not fully benefit from adding capacity, even though it may be beneficial to the system, and the benefits become "external" to the owners. Consumers and LSEs get benefits of a more reliable system but capacity owners do not get revenues reflecting that value. A free-rider problem is present in this situation because even if a consumer or LSE does not pay, he or she receives the benefits of a reliable system.

For similar reasons, reductions in demands during peak periods can also be viewed has having a positive externality. Jaffe and Felder (1996) argue that the magnitude of the reliability externality depends approximately on the product of the social cost of disruptions (VUE) and the probability of disruptions (P). Since this product is not small, Jaffe and Felder (1996) believe that the externality must be addressed by regulators.

An additional source of positive externalities may be electricity's pivotal infrastructure role in modern society (*ibid.*). The need for power to maintain communication networks and health systems, for example, could imply that the value of electricity to society is greater than just the consumer's willingness to pay for it. As a result, even if real-time prices accurately reflect VUE, they may understate the societal value of reliability and the need for capacity. Jaffe and Felder (1996) argue that this market failure can be addressed in the same ways that the informational market failures discussed in §2.2.2.1 can be corrected: by minimum installed capacity requirements or capacity payments. This should not be surprising, since the "information" and "externality" failures are two sides of the same "coin"—the basic failure of the market to provide prices for energy during capacity shortages that reflect its scarcity value.

2.2.2.3. Imperfect Competition. Imperfect competition is the term used for markets in which unilateral actions of an individual buyer or seller can affect market prices. This market failure can arise when some participants in the market are relatively large compared to the size of the market, or when transmission constraints restrict competition. This results in violation of the "no market power" assumption of the ideal spot market. Wolak (1997) explains that the desire of privately owned generation companies to maintain and attract shareholders implies that they will attempt to exploit any potential profit-making opportunities through their bidding behavior; thus, if they can manipulate price in their favor, they will attempt to do so.

There is a large literature on possible exercise of market power in electric energy markets. Several of the ways that energy prices can be manipulated are reviewed by Berry *et al.* (1999). But the literature has examined the relationship of market power to capacity decisions in only a few cases: short-term operating reserve markets (*e.g.*, unexplained price spikes in the California and ISO-NE markets) and capacity market manipulation in UK (Henney, 1998; Wolak, 1999).

Theoretically, producers might restrict capacity to keep energy or capacity prices high. As load approaches generating capacity, prices increase; by not making capacity available, generator owners can benefit from the higher prices and revenues.[13] This could occur if there are significant barriers to entry of new suppliers to the power pool. The net effect of this phenomenon is to decrease generation adequacy, unless the resulting higher prices induce more entry in the long run. The presence or absence of separate markets for capacity does not necessarily affect the ability of a large power producer to influence prices; we see no reason why the presence of market power should be viewed as a generic argument for or against capacity markets.

However, this is not to say that there may be particular circumstances in which exercise of market power might be more attractive for capacity markets than for energy markets, or vice versa. Singh and Jacobs (2000) point out, for instance, that PJM's dispersal of "Capacity Deficiency Rate" (CDR) receipts to generators that have surplus ICAP positions might remove some of the normal disincentives to capacity withholding. This is because a generator that decides to market its excess ICAP can still receive CDR revenue. As another example, if capacity markets have lower price elasticity of demand than energy markets, they may be more vulnerable to the use of market power (Stoft, 2000a). However, there is very little real-time pricing in PJM, so energy demands are also very inelastic, and the difference between energy and capacity elasticities may not matter much. However, the initial design of the ICAP market in New England eliminated what elasticity there may have been by automatically forcing LSEs with inadequate capacity to buy capacity in a daily ICAP auction. As a result, the demand bids were effectively perfectly inelastic, which invited suppliers of capacity to submit non-cost based bids in an often successful effort to push the price of ICAP credits to very high levels. Because of these problems, FERC allowed the New England ISO (ISO-NE) to abolish that daily auction, although not the ICAP credit system itself (FERC, 2000a).

Another way of exercising market power in the capacity market is for producers to declare capacity unavailable at key times. By doing this, the capacity in the pool is limited artificially and clearing prices in either the energy or operating reserves markets are increased. This was the case in the UK in which it was observed that suppliers declared plants unavailable to manipulate the reliability component of the day-ahead energy price. Subsequent to the price determination, but in time to participate in the market, the generators would put the plant on-line again to profit from increased prices (Wolak, 1999). This strategy was made possible by the relatively high market concentration in the intermediate-load and peaking plant market in the UK (Henney, 1998).[14,15] This type of exercise of market power has no apparent effect on generation adequacy.

---

[13] The PJM Market Monitoring Unit (PJM-Interconnection 2000d, p. 50) has investigated possible withholding of capacity during peaks. Although capacity availability was suspiciously low, they concluded that there was no evidence of strategic withholding.

[14] Henney (1998) notes that while in UK the Hirshman-Herfendahl index of market concentration was 1200 for base-load generation (indicating a relatively competitive market), the index for price-setting plants was of 3000, indicating a highly concentrated market based on US Department of Justice / Federal Trade Commission Merger Guidelines.

[15]Similarly high concentration in the market for operating reserves in California may explain the very high price spikes for reserve capacity observed there in the summer of 1998. Insufficient bids for reserves were received (even though total capacity was in ample supply), resulting in reserve prices that actually were well in excess of energy

In general, some of the ability of suppliers to manipulate prices would be mitigated if short-term demand was more elastic. Thus, the market power problem is due, in part, to the information failure identified earlier: consumers do not face or react to real-time prices. Other approaches to mitigating market power include structural measures (*e.g.,* breaking up large firms or encouraging entry), broadening the scope of markets (*e.g.,* allowing technically feasible substitution between different types of ancillary services), price caps or other regulatory instruments, and market monitoring.

2.2.2.4. Price Distortions Caused By Government Intervention and Market Design. This market failure refers to restrictions or adjustments in price imposed by market rules or regulators. One example in this context is price regulation or caps. Another is taxes or "uplifts" designed to pay for socialized costs (*e.g.*, system benefits charges or congestion costs). These interventions generally cause a divergence between the price paid by consumers and the price paid to producers. As a result, the marginal cost of production will not equal the marginal benefit of consumption (VUE).

Wolak (1997) notes that "market rules governing the operation of a restructured electricity market in combination with its market structure can have a substantial impact on behavior of market clearing prices." Market structure and rules governing the operation of the electricity industry are not a direct result of independent actions by market participants (generators, customers, and retailers), but the outcome of a deliberate government policy to restructure the industry. Wolak then states it is "a misnomer to call these markets competitive." He invites us to think of market structuring as another form of regulation with the goal of greater economic efficiency in supply of electricity. It is clear that the operating rules in the market will be continuously modified to correct actual and perceived problems.

But modifications of market rules designed to correct one set of problems can exacerbate other market failures. An example is when an artificial price limit is set. Price caps have been called for in the wake of the 1998 Midwest and 2000 California energy price spikes (Hirst *et al.*, 1999; Michaels, 1999), and are imposed upon operating reserves and energy markets in New England and California. PJM has set an energy price cap of 1000$/MWh. However, if a price cap constrains price to be below what people are willing to pay for it, consumers are further isolated from the opportunity cost of energy use—prompting an information failure that can decrease allocative efficiency.

As another example of an allocative inefficiency, Stoft (2000a) points out that a combination of an energy price cap and a maximum penalty for ICAP credit shortfall can encourage PJM power producers to divert capacity to other regions whose energy price is not capped during periods of price spikes. This is because it can be more profitable to accept the penalty and sell the power at an uncapped price elsewhere than to keep the capacity within PJM and sell energy there at the capped price.[16] Finally, long run productive efficiency can also be affected if price caps discourage needed capacity additions.

As pointed out in §2.2.2.1, changing policy and market designs adds to uncertainty, and this might discourage investment, as Hirst *et al.* (1999) point out. This uncertainty could also impact

---

prices. To deal with this problem, California requested and received authority from FERC to place caps on reserve prices.

[16] PJM does possess certain rights to recall capacity back into its market; see Appendix III.

the demand side. Uncertainties concerning the distribution of prices and how shifting designs might affect it could make potential participants wary of real-time pricing programs. This will compound the existing uncertainty concerning how customers will react to such programs (NERC, 1998, cited in Hirst *et al.*, 1999).

A final effect of policy upon market prices is the imposition of reliability constraints. Of course, such constraints (when translated into, *e.g.*, installed capacity requirements) are usually imposed to correct some of the other market failures mentioned above. However, Graves and Read (1998) and Jaffe and Felder (1996) point out that if the reliability standard is poorly chosen (one that is well below or above the level at which marginal benefits and costs are equated), then social net benefits can actually be decreased relative to a situation with no intervention. Basing such standards upon traditional industry rules of thumb (such as "one day in 10 years") rather than a thorough understanding of the VUE increases the likelihood of that unhappy outcome.

3. CORRECTING FAILURES IN THE MARKET FOR RELIABILITY: ISSUES AND SUPPLY-SIDE PROPOSALS

Restructured power markets around the globe have taken a variety of approaches to ensure adequacy of generation capacity. One approach is to correct the basic market failures that prevent consumers from expressing their WTP for energy and reliability in real time. By increasing demand-side price elasticity and responsiveness, price signals can better reflect valuations of reliability. Examples of this approach include:

1. Expansion of real-time pricing and interruptible load programs;
2. Coincident peak pricing programs in which consumers are guaranteed to face high real-time prices only a limited number of hours per year (Uhr, 2001). These hours would be chosen by the load serving entity to coincide (to the extent possible) with times of peak prices. An experiment at Seattle City Light showed that this approach would be attractive to large customers.
3. Capacity subscriptions (Doorman, 2001), inspired by theoretical work by Panzar and Sibley (1978). Consumers would buy a “fuse” that limits them to certain demand levels during times of capacity shortages.

The other approach addresses the supply side: providing artificial incentives for generators to add sufficient capacity. There are numerous supply-side proposals, each having the goal of correcting market failures that would prevent the spot energy market by itself from being able to achieve the optimal level of capacity. Jaffe and Felder (1996) pose the basic question that market designers must ask: Should a capacity market be set up to give an explicit value placed on capacity, or should the level of capacity be driven by expectations about the spot price for electricity (and perhaps operating reserves), as reflected in prices in forward markets? Where a separate market for capacity is created, either of two basic approaches is adopted: a price-based capacity system or a quantity-based capacity system. It is rare to find a pure energy spot price approach; however, some systems such as California have allowed energy and operating reserves markets together to provide the only incentives for capacity additions. Because the operating reserves market is driven by a regulatory constraint on the amount of such reserves, rather than by customer valuation of reliability, that approach is philosophically related to the quantity-based installed capacity system—and is so treated here. Thus, since there is no market in which direct customer valuation of reliability determines capacity additions, several authors claim that a true market based solution to reliability remains to be constructed (Stoft 2000b; Henney 1998).

In theory, if the optimal level of capacity can be identified, then either a price-based or a quantity-based (either installed capacity or operating reserves) system can be used to achieve it (Jaffe and Felder, 1996; Stoft, 2000b). We illustrate this point in §4 for a stylized representation of the PJM system. There, under certain assumptions, we show that with the correct prices, capacity reserve requirements, or operating reserve requirements, the socially optimal amount and mix of peaking and baseload capacity will result. Thus, any of these mechanisms can correct the market failure that energy prices do not reflect customer WTP for reliability. This is accomplished by implicitly or explicitly providing extra payments for capacity so that the private returns to investors resulting from building capacity align with the social benefits of providing that capacity.

This section summarizes basic features of the price-based and quantity-based systems (§§3.1, 3.2), and discusses how they address the market failures just identified. Then we summarize some general advantages and disadvantages of each system (§3.3), focusing on ICAP-type systems.

3.1. Price – Based Systems

One approach to correcting market failures concerning capacity is for the ISO to pay generators directly for capacity, which increases the profitability of new capacity investments. This approach has been followed in several countries with different results, notably in the UK, Argentina, and Spain. The payments can take two forms. The first is a payment for installed capacity separate from payments for energy, as in Argentina. The system there has been abolished as of March 2000. In Spain, the capacity payments are similar to stranded investment compensation (Oren, 2000). The second form of capacity payment is an uplift in the energy payment that depends on the state of the system and the capacity availability, as in the UK system before March 2001.

The use of a separate capacity payment is a logical continuation of the practice under traditional utility regulation of calculating a capacity component of revenue requirements based on the book value of installed capacity and an allowed rate of return (Graves and Read, 1998). This practice permitted utilities to earn a return on generation investment needed to meet potential peak demands under the universal service goal—even if that plant rarely if ever actually generated power.

The calculations for the capacity payment differ from market to market. Frequently, the payment is based on the cost of a peaking plant, such as a combustion turbine.[17] In Argentina, the capacity price was fixed and determined by the Central Dispatch Authority. The capacity price allowed generators to make the transition from a system of government-owned utilities to a competitive private market. Because the capacity price did not depend on location, some plants that could not deliver their output to where it was needed because of transmission constraints nevertheless received payments.

In contrast, prior to the New Electricity Trading Arrangements implemented in March 2001, the UK had a sophisticated real-time system that augmented the energy price with a capacity adder under certain conditions. The adder was calculated as the product of the each half-hour's LOLP times an assumed VUE. This calculation paid generators a price above the marginal cost of energy, promoting the installation of capacity. The VUE was imposed by the market operator rather than resulting from the interaction of customers and suppliers. As Henney (1998) observes, such calculation of prices by a central organization is fundamentally different than attempting to allow consumer WTP for reliability to be revealed via the market.[18]

Such payments to owners of installed capacity are directed at correcting the information and externality market failures described in §2.2 in two ways. First, they increase prices received by

[17] Oren (2000) presents a rationale for this procedure.

[18] Market mechanisms could be used to reveal customer willingness to pay for reliability if, as pointed out elsewhere in this report, customers see and react to spot prices in real time. Another market mechanism for allowing customers to express WTP include the Chao-Peck priority pricing system (1998), in which (1) power is differentiated by its reliability and (2) customers self-sort according to their WTP for higher levels of reliability.

generators so that the reliability benefit received by consumers are appropriately considered by generators. Second, these payments can dampen energy price volatility (especially if coupled with energy price caps), thus lowering the uncertainty associated with investments. Graves and Read (1998) and Singh and Jacobs (2000) explain that capacity could be interpreted as a call option on energy, and in this role, it could reduce volatility. If capacity payments result in more capacity being added, there will be fewer times when prices spike upwards as a result of shortages. Given the political consequences of the price spike episodes elsewhere in the nation over the last three years, the decreased price volatility that results from a combination of an ICAP market and caps on energy prices may enhance the public acceptability of market restructuring.

But price volatility is a result of a real variation in the scarcity of resources, so dampening it artificially with the energy market price caps that accompany ICAP might be undesirable. Singh and Jacobs (2000) argue that price volatility creates opportunities for retail energy providers and market traders to provide price insurance via price futures or options in the wholesale market. These market participants could offer the consumer a greater choice by serving customers that are willing to accept greater uncertainty in exchange for lower expected energy prices. Singh and Jacobs argue that volatility will increase the demand for and availability of long term financial contracts, ensuring a more predictable revenue flow for generators. Price volatility also generates incentives to create price sensitive demand, which is one of the key failures in existing power markets. Stoft (2000a) points out that measures to artificially reduce price volatility will discourage economic demand adjustments during those times when capacity is short.

3.2. Quantity – Based Systems

Capacity payments are one way to deal the failure of energy markets to provide sufficient installed capacity; another is the creation of a separate market for capacity. There are several systems that can be designed in a continuum, ranging from an extreme pure reserves market system to a pure Installed Capacity Requirement based system. A mixture of both could be the proposed design for the Available Capacity Market for PJM-WEST (PJM-Interconnection, 2001). Three quantity-based systems are discussed here. Two are based on a (relatively) long run market for installed capacity—the ICAP system (§3.2.1) and a mandatory call-option proposal (§3.2.3). The other quantity-based system is a short-run market for operating reserves that creates additional incentives to build capacity (§3.2.2).

3.2.1. Installed Capacity Requirements and Credit Markets

Jaffe and Fedler (1996) point out that under a fixed capacity payment scheme like Argentina's, the amount of capacity that the market will provide cannot be predicted exactly. An alternative is a quantity-based system, such as an installed capacity requirement with tradable credits. In the ICAP system, the ISO would:

- define how much capacity is needed, based on the system load and the desired reliability level;
- allocate responsibility for providing that capacity among market participants;
- allow the participants to trade capacity credits so that capacity is provided in a least cost manner; and

- establish a system of penalties for non-compliance.

The logic of capacity markets is different from the priced-based systems in that the generators are not paid a fixed amount of money for capacity based on an assumed VUE or cost of peaking capacity. Instead, the value of capacity results from the interaction of buyers and sellers in the capacity market; if there is a surplus of capacity, then the value of capacity credits will be low. A shortage of capacity will, in contrast, result in a higher value. An owner of capacity realizes this value either by selling its capacity credits, or using them to fulfill its own obligation, if it is a LSE. In the long run, the price of capacity credits should fluctuate around the cost of peaking capacity (Graves and Read, 1998).

The best known example of capacity markets is the PJM ICAP system, described in more detail in Appendices II-IV of this report. ISO-NE and NYISO also have ICAP markets (Rau, 1999; Sotkiewicz, 2000), although ISO-NE has modified its version (Cramton and Lien, 2000). In the case of PJM, the ISO calculates the amount of required capacity required in each year considering the summer peak load, available load controls, and the forced outage rate experience of existing generators. PJM assumes that the responsibility for ensuring generation adequacy should lie with the entities responsible for serving load, the LSEs; consequently, each LSE must secure capacity or capacity credits sufficient to cover its contribution to the coincident summer peak. This responsibility is enforced by imposing an ICAP obligation upon each LSE. LSEs can obtain or sell capacity credits in bilateral deals, or through daily and monthly ICAP markets set up by PJM.[19] An important feature of this market is that the ICAP price is effectively capped by the financial penalty imposed if a LSE fails to have sufficient credits. If this penalty is low enough, it could wind up setting the price of capacity at some times (Appendix III), in which case the PJM system effectively becomes a price-based rather than quantity-based capacity market (Singh and Jacobs, 2000; Stoft, 2000a).

Just like the price-based system, the quantity-based system addresses the informational and externality market failures. The failure of energy prices to reflect consumers' VUE (or, equivalently, the externality value of reliability) is dealt with by setting the required level of capacity to a level that maximizes the estimated social net benefits. Under pure competition, the price of capacity credits will then rise high enough to motivate sufficient additions; theory says that the resulting total returns to capacity investment should be the same as if the energy price correctly included the externality (§4).

Like the price-based system, a quantity-based system can also help depress the volatility of prices by (1) decreasing the frequency of shortage; and (2) allowing the imposition of energy price caps to occur while still motivating construction of sufficient capacity. Reducing this risk might encourage construction of more capacity. On the other hand, uncertainties in whether the design of the capacity market will be changed in PJM and elsewhere add risk, perhaps discouraging investment. Furthermore, ICAP prices themselves can be volatile, as recent experience in PJM shows; thus, the aggregate risk faced by LSEs might not be less than under a pure energy market.

[19] At PJM, the market creates contracts between the participants that have one day, one month, or six-month duration. In contrast, the NYISO runs seasonal 6-month capacity obligations markets.

3.2.2. Operating Reserves Markets

Instead of creating a market for long-term installed capacity, markets could instead be created for short-term operating reserves. Assuming that reserves prices would be highest during peak periods, this system would reward capacity that is available when it is most likely to be needed, as opposed to just rewarding "iron in the ground." Markets can be created for several types of reserves simultaneously, such as for regulation reserve and for reserves that can be made available within 10 and 30 minutes. ISO-NE and California have such markets, in which the ISO purchases a predetermined amount of reserves from the lowest bidders. The amounts of reserves required are based on reliability studies that consider various short-term demand and equipment outage contingencies.[20,21]

The resulting payments for reserves provide a revenue stream for plants that can be used in emergency or shortage situations, even if they never actually generate power. If the amounts of

---

[20] The NY market has operating reserves markets for several sub regions of the state. In contrast, PJM has no reserves market. It uses a nonmarket process to determine who supplies reserves based on a least-cost solution from its day-ahead energy market model. Suppliers of reserves receive payments only if their costs would otherwise be uncompensated. In PJM, the amount of regulation required is approximately 1.5% of peak demand, while the amount of operating reserves vary from 2000-3000 MW (Stephen Stoft, personal communication, 2000).

However, PJM is considering operating reserves systems as alternatives to ICAP. In particular, the Operating Capacity Obligation - OPCAP - is a modification under study in the FAWG at PJM.(Bhavaraju, 2000). The system that would replace ICAP would consist of a mix of ICAP ready to start up in 30 minutes, capacity with 30 minutes operating reserves characteristics, and load management response available within 30 minutes. This mix is equivalent to the operating reserves concept of this section. OPCAP obligations would be based on estimates of loads for specific periods of the next day supplied by the LSEs. The system is designed so its performance complies with the current MAAC adequacy criterion of a Loss of Load Expectation of 1 day in ten years. Trading through bilateral or PJM auction markets will satisfy the OPCAP requirements. If a LSE is in deficit, PJM will bill the non-complying LSEs for the cost of the purchase. This concept of OPCAP is different from the market that ISO-NE eliminated recently (Cramton and Lien, 2000).

Meanwhile, the New England ISO is presently supporting an operating reserves market as an alternative to its ICAP market. In particular, they argue that the development of reserve markets based on demand curves as described in the March 31 Filing (ISO-NE Documents 2000) will substantially improve the flaws of the current reserve markets. The demand curve yields a price for reserves that depends on the amount of reserves available; if reserves are in short supply, the price is higher. This type of reserve market was first proposed by Stoft (2000b). Its advantage is that it avoids the “bang-bang” type of reserve prices that the operating reserves market analyzed in §4 exhibits, in which the price is either zero or the maximum level.

ISO-NE has formed a working group within its Markets Committee to promptly formulate revisions to the reserve markets including forward purchases of reserves. The ISO argues that long-term options for reserves are real products with real costs associated with their delivery. It is hoped that they will serve as a basis for an efficient market for capacity, which will not be subject to the pricing abnormalities associated with the current reserve and capacity markets.

[21] Just as the difference between price-based and quantity-based systems is blurred in reality, so can the distinction between installed capacity and operating reserves markets. The PJM ICAP system imposes a daily obligation on LSEs. Because of the threat of capacity withdrawals during peak periods (see §3.3.3) and because of concerns about poor availability during those periods, the PJM Market Monitoring Unit (PJM-Interconnection 2000d, p. 58) has proposed that higher penalties be assessed for capacity withdrawals at those times and that capacity that is unavailable during peaks get penalized more. This focus on availability during peak periods moves the ICAP market in the direction of becoming, in effect, a system of rewarding capacity that can provide operating reserves or energy on a daily basis rather than "iron in the ground" on an annual basis.

operating reserves that are bought by the ISO are appropriately set, additional capacity will need to be installed, and the desired level of long-run adequacy can in theory be achieved (§4).

Singh and Jacobs (2000) argue that this incentive is sufficient to correct the failures of a pure energy market, and no ICAP-type installed capacity markets are needed. The basic reason is the same as given in §3.2.1 for installed capacity markets: if operating reserve requirements are set to socially optimal levels, returns on generation investments will increase because of reserve revenues. (Of course, if the ISO assumes an incorrect value for VUE or chooses the wrong level of operating reserves, then the amount of installed capacity will not be socially optimal.)

In spite of the apparent differences that exist between an ICAP requirement and the operating reserve auctions, they are conceptually linked. In both cases, the ISO determines what reserve margin is needed to operate the system safely. In the PJM ICAP system, for instance, a reliability assessment is undertaken for a one-year period (see Appendix II) and a required installed reserve margin is imposed on LSEs. In the operating reserve system, the ISO sets a short-term reserve level adequate to cover anticipated demand and equipment contingencies. Although these seem like entirely separate processes, the long-run reserve analysis can be basically viewed as an average over all the short run conditions. Given a long-run installed capacity reserve requirement and the short-term operating reserve limit, standard capacity reliability methods can be used to calculate the expected hours per year that the operating reserve limit will be violated. The calculation can also go the other way: given the operating reserve standard and number of hours per year that it is allowed to be violated, the implied long-run installed reserves can be calculated. As a result, high prices for reserves during times of reserve shortages can be translated into an equivalent price for ICAP. Each has a similar positive benefit on capacity investment decisions, as will be shown in §4.

In addition to affecting incentives for building capacity, operating reserve markets will decrease energy price volatility (Stoft, 2000b) under the following assumptions:

- either additional capacity is added so that shortages occur less often, or energy prices are capped at a lower level than would have been otherwise; and
- the ISO caps the price it pays for operating reserves prices at a moderate level (say, several hundred $/MW/hour), so that if there are insufficient reserves, reserve prices rise to that level. At such times, energy prices will be equal to the reserve price, plus the marginal cost of fuel, under the assumption that power plants effectively arbitrage between the markets by being able to choose whether they provide reserves or energy.[22]

Under these assumptions, instead of high energy price spikes a few hours per year, there will instead be moderately high energy prices for a larger number of hours per year (those hours in which there are insufficient reserves). Volatility will be less, as measured by, for instance, the standard deviation of prices within a period. This issue is explored in the analyses of §4.

[22] The ability of a unit to sell in either the reserves or energy market means that, at the margin, there is some unit that will be indifferent between the two. It can then either sell its capacity as reserves in the reserve market at the reserve market price (incurring little or no fuel cost in doing so), or generate and sell power in the energy market. For the unit to be indifferent, the prices in the two markets can therefore only differ by the cost of fuel for this marginal unit (see §4).

3.2.3. Mandatory Call Options

This modification of the ICAP market has been proposed by Oren (2000) and Vázquez *et al.* (2001), promoted within PJM by Singh (2000), and mentioned favorably in a FERC Staff document (FERC, 2001). The purpose of the proposal is to create an availability product with real value in the market, unlike ICAP, which is essentially an artificial product created by regulators with no inherent value. Like ICAP, the proposal would have the ISO require that all LSEs own or sign contracts for "capacity" in proportion to their projected peak demand (adjusted for load diversity), with credit given for capacity they own and load control programs. However, this "capacity" requirement would be in the form of call options with a relatively high strike price (*e.g.*, $999/MWh) rather than ICAP credits. If called, the seller of the option could satisfy their obligation to the LSE in either of two ways: (1) by physically providing the contracted energy at the strike price or (2) by paying the difference between the unrestricted spot price and the strike price. If an ISO desires, the seller of such a call option could be required to own or contract for capacity to back up the options it sells.

The authors of this proposal suggest that the cap on the energy spot price would be removed, and that the strike price could be set equal to the former price cap (although variants of this proposal would allow LSEs to choose their own strike price). As a result, the LSEs would be protected from energy prices exceeding the cap, as in the present PJM system. However, as discussed further in §3.3.2.3 below, generators would have a more appropriate incentive to improve plant availability during peak periods in order to protect themselves from having to make payments to option holders during price spikes. Furthermore, incentives would disappear for delisting ICAP and diverting energy production to neighboring areas not subject to spot price caps. Finally, the option has a real value that can be assessed using option valuation techniques, unlike ICAP.

As an example of how this proposal would work, consider the following. Say that the PJM price rises to $1500, which is over the strike price, of say, $999/MWh. A PJM LSE with this contract would then exercise the call option, and the supplier could provide the energy at $999/MWh if its unit is available. Alternatively, the seller would make a payment of $501 ($1500-$999); the LSE could then buy energy on the spot market, paying a net price of $999 (= the market price of $1500 minus the payment of $501 received from the exercise of the call option). If all LSEs in PJM had sufficient call options to cover their demands, then they could all meet their demands at a cost of no more than $999.

In theory, if market participants are risk neutral, the equilibrium price for such an option should approximate the probability weighted average of the positive difference between the energy price and the strike price over the year. For instance, if the spot price is expected to be below the strike price of $999 for 99% of the year, while for the rest of the time it is expected to reach $1499, then this probability weighted difference is 0.01x($1499-$999)x8760 hr/year = $43,800 per year. It turns out that this is the same as the equilibrium price of ICAP in the pure competition model of §4. If desired, the ISO could cap the price of the option at a certain level (*e.g.*, some multiple of the cost of a combustion turbine) by allowing LSEs to pay a penalty equal to that amount if capacity shortages arise.[23]

[23] The market price of such call options would in all likelihood be significantly higher than calculated here, given the risk averse attitudes of market participants along with uncertainty as to whether a chronic California-type shortage could occur.

Oren (2000) suggests that purchasers of the call option determine what strike price they want and what contract duration they prefer, rather than have it set by the ISO. For instance, a LSE that prefers to bear more risk would sign contracts with higher strike prices (say, $2000 rather than $1000/MWh) and shorter durations (*e.g.*, 1 month instead of 1 year). Oren (2000) points out that as the contract duration shrinks, the price of the call option will fluctuate more until it starts behaving like a spot market for capacity. However, for the sake of smoothness of transition, we make the following recommendation. If PJM decides to switch from an ICAP system to the mandatory call option approach, the required option should at least initially be required to have the same or lower strike price as the current energy cap ($999), and the duration be relatively lengthy (one season or one year).

It has been suggested that the options acquired under this proposal be required to be backed up by physical capacity (FERC, 2001). In that case, the distinction between ICAP and mandatory options is blurred, as a distinguishing feature of ICAP is its link to physical generation plant.

Although there are theoretical advantages to this proposal, concerns that have been raised about several practical aspects (P. Lalor, Commonwealth Power Corp., personal communication, 2001). Some of these include questions about the type of option to be required (European, American, or other), the time frame in which it would be exercised (based on day-ahead prices, hourly prices, or balancing prices), and the time frame for which the contracts would be effective (week, month, or year?). Another important question is whether enough willing suppliers of such options could be found, given the risks involved and the desire by suppliers to keep open the option of supplying the balancing market. However, it should be noted that many of these practical issues have apparently been successfully dealt with in a mandatory call option proposal nearing implementation in Colombia (Vázquez *et al*., 2001).

3.3. Advantages And Disadvantages of Alternative Capacity Market Designs

This section summarizes selected pros and cons of the various systems based both on actual experience where they have been implemented, and on theoretical discussions in the literature. We emphasize the criticisms that have been directed at the ICAP system. We should note before reviewing the relative advantages of different systems that, in practice, the distinction between price-based and quantity-based systems is not as clear cut as our discussion might seem to indicate. For example, as pointed out earlier, quantity-based markets for installed capacity (such as the PJM ICAP system) can allow market participants to pay a penalty if they have insufficient capacity credits. If the penalty is low enough, its effect on decision making can be similar to systems based on a payment for capacity. As another example, the UK system (in which price premiums are paid for energy during periods of shortages) has a similar effect on energy prices as an operating reserve market. This is because the high value that an ISO pays for reserves in times of shortage can cause energy prices to rise in parallel.

It should be noted that both the price-based and quantity-based systems suffer from the problem that, at their heart, they are driven by administrative decisions (*e.g.*, about assumed values of VUE, capacity payment rates and penalties, or required margins), and not by market information that reveals what consumers are really willing to pay for reliability. Either system can result in significant distortions if the administrative decisions are ill-informed—either by using a penalty that greatly deviates from the VUE or cost of capacity, or by imposing a margin that is far from the socially optimal level. (However, as will be noted below, it has been argued that the cost of errors may be less for quantity-based systems than price-based systems.)

3.3.1. *Price-Based Systems Advantages and Disadvantages*

Three issues have been raised concerning price-based systems: the burden upon the ISO's budget of having to make payments; the simplicity or complexity of the systems; and the difficulty of predicting how much capacity will result from such systems.

The price-based system suffers from the practical problem that the ISO will have to make payments for capacity that, in turn, have to be recovered from consumers, usually in the form of an uplift. If this uplift is not assigned to peak demands, then there will be too little incentive to reduce consumption at those times. Furthermore, if an uplift is applied to pool transactions, then market participants will have an incentive to turn to bilateral transactions. In contrast, quantity-based systems based on tradable capacity credits do not cost the ISO anything—and indeed could be a source of revenue from penalties paid by participants who are short of credits.[24] (However, if the quantity system is designed as an auction for capacity in which the ISO buys capacity from winning bidders, then large ISO expenditures are involved. An example of such a system is the operating reserve auction system described above.)

Price-based systems in which a simple fee is paid for installed capacity are relatively simple to administer. However, price-based systems in which the payment depends on system conditions—as in the UK before this year—are complex (Henney, 1998). Prices were calculated by the UK ISO using an opaque set of rules. Many of the rules were gameable, and the costs of different services, as ensuring generation reliability, short-run costs of transmission, and ancillary services for security, are transferred to customers as "uplifts." Initially it was estimated that the uplift would be 2% of the traded value, but it reached 10% in 1993. In part because of this complexity, this system was abandoned in 2001.

The final issue raised about price-based systems concerns the effects of fees that ignore market conditions upon generation adequacy. Predicting the capacity that payments will elicit from the market—and the resulting system reliability—is in theory a difficult problem (Jaffe and Fedler, 1996; Ruff, 1999). If payments are fixed on a $/MW basis, and if the supply of capacity is very price responsive (elastic), then a small change in the payment could have a significant effect on installed capacity. Underestimating the payment needed to elicit additions could result in too little capacity, while setting the payment too high can result in far too much capacity.[25] In contrast, a quantity-based system is more likely to yield a predictable amount of capacity additions, which especially in the present uncertain political climate is a distinct advantage of the quantity system.

---

[24] PJM redistributes the penalty payments it receives to generation owners with capacity surplus.

[25] The experience in Argentina illustrates this point. During this decade, the fee has led the system to install much more capacity than it was needed to meet demand, especially gas turbines at the head of the gas wells, without properly considering transmission constraints. It has turned out that these constraints are the cause of the blackouts suffered by the country in the last two years (F.C. Rey, Comisión Nacional de Energia Atómica, Argentina, personal communication, Jan. 2000).

3.3.2. Quantity-Based Systems (ICAP) Advantages and Disadvantages

In this section, we concentrate on ICAP as a quantity-based system to analyze the advantages and criticisms to its implementation. The advantages and disadvantages of operating reserves markets have already been summarized above (§3.2.2), but will also be discussed below where that type of market has been proposed to correct problems of ICAP. Brief mention will also be made of the mandatory call option system of §3.2.3.

There are two major advantages to installed capacity markets. The first is that there is more assurance that minimum capacity and reliability targets will be achieved, relative to price-based systems (assuming that penalties for non-compliance are large enough). The second is that installed capacity market are subject to less volatility, in theory, than operating reserves markets, and may be less prone to manipulation. However, installed capacity markets are also argued to have many disadvantages. For instance, they can affect ("distort", in some people's opinion) energy prices; they are alleged to increase the cost of power unnecessarily; they have a less direct relationship to system reliability than operating reserves; and particular aspects of their design may make them vulnerable to "reserve stealing" by neighboring power systems. The last issue is the focus of a separate subsection. The other pros and cons are summarized below. The criticisms are organized in two lists, one grouping the critics that are inherent to the concept of ICAP market as a solution, and the second, those that arise depending on the actual market design.

3.3.2.1. Advantages. Jaffe and Felder (1996) support the existence of ICAP markets because of their predictability: given a sufficiently high penalty for noncompliance, the market will provide at least as much capacity as is required by the ISO.[26] In contrast, if a fixed payment is made for capacity, too much or too little might be provided. Their analysis shows that if there is a threshold above which the marginal benefit of capacity is very low, and above which is very high, then the social cost that results from setting the wrong installed reserve requirement (in an ICAP-type system) is likely to be smaller than from setting the wrong payment in a price-based system.

Installed capacity systems may also have a predictability advantage over operating reserves systems. Operating reserves markets are often thin (few suppliers), and as a result prices can be very volatile because of attempts to manipulate the market (Sheffrin, 1999). For this reason, New England and California received FERC permission to cap their reserves prices. Installed capacity markets, on the other hand, are likely to evolve over a longer period of time—particularly if they involve long-term (*e.g.*, 6 month) commitments—and involve more players. As a result, prices should be more stable and reflect greater competition. (However, PJM has experienced wide variations in daily ICAP prices, so this advantage may not be a real one. Further, the ability of capacity owners to pull their capacity out of the PJM ICAP system at short notice also erodes the predictability advantage.) It is important to note that ICAP requirements are forecast three years in advance by PJM, which provides some "early warning" of the need for capacity additions.[27]

---

[26] What constitutes a sufficient penalty is an important issue. Clearly, the penalty has not been high enough in PJM; see Appendix III.

[27] However, a liquid futures market for energy or operating reserves could play the same role in non-ICAP systems.

Stacked against the advantage of predictability are several alleged disadvantages, described in detail below. The disadvantages are divided into two groups: those that are inherent to the ICAP concept (§3.3.3.2), and those that depend on particulars of implementation (§3.3.3.3).

3.3.2.2 Disadvantages Argued to be Inherent to the Concept of ICAP. The disadvantages that critics of the ICAP system argue are inherent to the basic ICAP idea include: it unnecessarily complicates the market and thereby increases uncertainty; ICAP distorts energy prices, ICAP distorts entry/exit incentives, and ICAP has only an indirect relation to system reliability. These are discussed in turn below. Where conclusions from the quantitative analysis of §4 are relevant, they are summarized.

ICAP Complicates the Market. Hirst *et al.* (1999) object that capacity markets create costs and uncertainties associated with rebundling two products—capacity and energy—that they argue do not need to be separated in the first place. Having to deal with two commodities rather than one makes transactions more complicated than they need to be. Uncertainties are introduced because additional gaming opportunities might be introduced, and market institutions may be changed in unforeseen ways.

ICAP Alters Energy Prices. An objection mentioned by Hirst *et al.* (1999) is that existence of a capacity market will alter the level or distribution over time of energy prices. In §4, we show that this can indeed be the case, although the resulting changes in energy price distributions are not necessarily inefficient (indeed, they could be more efficient). In general, inefficiencies result if short-term dispatch is distorted, inefficient maintenance decisions are made, or, as discussed later, inefficient additions or retirements of generation assets occur.

As an example of this criticism, in a filing presented by the ISO New England (2000b) to FERC, they argued that if the requirement for installed reserves is set higher or lower than what it is economically efficient, the spot market price could be then be below or above, respectively, the efficient level. But this is not only a disadvantage for ICAP, but for other reserves requirements in general, whether provided in an ICAP mechanism or by a operating reserves ancillary market. Any reserve requirement can be set to a non-optimal level. Prices could be distorted, but as we illustrate in §4, if there is no consumer short-run response to prices, then, in the long run, the same capacity decisions can result from a pure energy market and separate energy and capacity markets (see also Stoft, 2000b). The total costs paid by consumers should be the same in the long run, although their distribution over the hours of the year (reflected in the volatility of prices) will differ.

Besides the potential efficiency drawbacks of distorting energy prices, it has also been argued that these distortions hurt consumers by raising the overall cost of power. For instance, Hanger *et al.* (2000) state that PJM's ICAP market "unnecessarily add(s) from 0.4 to 1.8 cents per kWh to the cost of serving a residential customer." These costs are consistent with the average 1999 ICAP credit cost of $53/MW/day (Market Monitoring Unit, PJM-Interconnection, 2000g). However, as §4 shows, in the long run, whether or not a capacity market is used should not affect the total cost of power to consumers. In that analysis, it turns out that the cost of credits is compensated for by lower energy costs. Nevertheless, this moderating effect on energy prices was claimed not to exist in PJM (McCormick, 2000). This moderating effect should be possible to observe empirically through comparison of energy prices in PJM and a neighboring region without ICAP (*i.e.* ,ECAR).

The conclusion of §4 that an ICAP market will not add to the total cost of power supply assumes, among other things, an absence of market power. However, there are claims that there is a scarce ICAP credit supply (McCormick, 2000) due to market concentration. Events where suspicious ICAP prices have been observed have been investigated by the PJM Market Monitoring Unit (PJM-Interconnection, 2000d). In particular, during June 2000 the ICAP prices spiked, but the effect was explained as being due to the increasing energy prices in neighboring markets and capacity delisting, not by market manipulation (see the discussion in §3.3.2.3, *infra*). Appendix V, below, summarizes the events of that month.[28]

Nevertheless, the existence of significant prices for ICAP credits even when capacity is in surplus and PJM mandatory market rules forced that capacity onto the market suggests that market power may be a problem. Because there are some O&M costs associated with keeping installed capacity ready for operation, the ICAP credits should not have a zero price; but the prices have been well above that level for much of the past year. The high ICAP prices that occurred in the New England ISO market are attributable to particular design flaws in that market, discussed in §3.3.2.3, below.

One possible explanation for the significant prices for PJM ICAP credits are their opportunity costs. In particular, the analysis of §4 implies that the fact that PJM energy prices are capped and that prices were sometimes higher elsewhere would mean that there is a shadow (opportunity) cost of capacity committed to PJM—the revenues foregone elsewhere. In equilibrium, it could well be that that this shadow price integrates, in some fashion, the price difference between PJM and other markets. This should be examined empirically.

The argument that an ICAP market distorts energy prices (and thereby artificially increases the total price of power) is often made by asserting that consumers buy capacity twice: (1) as firm energy and (2) as ICAP credits (McCormick, 2000). This is not obviously true. Firm energy is negotiated through contracts or can be obtained by paying the spot price at all times. Since ICAP can serve to lower the average spot price (under the price cap and perfect competition assumptions of §4), then the price of firm contracts that exclude ICAP credits should also decrease.

However, as pointed out earlier in this report, consumers could be paying twice if ICAP they have paid for is delisted and then diverted to other markets during times when energy prices are higher elsewhere. This means that the capacity has been paid for, but it is not available when it is most valuable and the system will have to pay a premium for power to replace it during those periods.

<u>*ICAP Alters Entry/Exit Decisions.*</u> To qualify as a producer in PJM, a company has to obtain approval from the ISO, based on capacity ratings of the plant and a forced outage index. The producer has to submit performance reports for their plants to the PJM generator data base ("GADS"). This information feeds the procedure for being approved as an ICAP resource and as

[28]The precise way in which the amount of capacity credit a given resource is assigned is an issue that could conceivably distort energy prices if done in an arbitrary manner. A complaint has been made, for instance, that the definitions of capacity ICAP resource ratings do not reflect winter maximum capabilities, which tend to be higher (McCormick, 2000). This is done because ICAP requirements for the pool are calculated based on the summer peak. It has been suggested that the use of the more conservative capabilities may exaggerate how much capacity is required; this can depress spot prices by promoting more than the economic efficient level of capacity.

a bidder in the market. Hirst *et al.* (1999) (citing Henney, 1998) says that the filing requirements associated with a capacity market along with the costs of participating in the market may add an additional obstacle to entry by new producers. On the other hand, the existence of a capacity market might result in a steadier (if not necessarily larger) stream of revenues. Given the large number of proposals for new plants in the PJM region, it does not seem likely that entry has been discouraged by the complexity of filing for ICAP status and participating in the market.

Another concern that has been expressed about ICAP's effects upon entry and exit of plants is about whether the existence of the market somehow discourages retirement of old, inefficient units with high emissions rates. If ICAP indeed lengthens the life of such units, then ICAP could be indirectly leading to higher emissions of air pollutants. For instance, the ISO-NE said in a presentation to FERC that "(t)he Installed Capability market provides a payment that distorts the signals generated by the energy and ancillary service markets. By subsidizing the continued operation of inefficient units, and maintaining reserve capacity at or above the required reserve margin, spot prices are reduced below their efficient level" (ISO-NE, 2000b). Another argument says that without the distortion of the energy prices, a generator bidding its variable costs in the energy market and clearing near the margin (setting or nearly setting the clearing price) might not, absent ICAP payments earn revenues sufficient to keep it in business. Without the ICAP payment, the argument goes, the plant would either shut down or submit higher energy bids. If those bids no longer cleared it would shut down.[29]

However, the analyses in §4 imply that there is no such distortion in a purely competitive market. In particular, under the assumptions made in that section, the profitability of any existing plant will be the same for any of the following systems: ICAP, operating reserves markets, capacity payment, and pure spot market. The key assumption is that the ICAP, operating reserves requirements, and capacity payment are set at a level that will result in the same amount of total available capacity as the ideal spot market. It turns out that any plant whose running costs are no more than that of the peaking plant (combustion turbine) in that analysis will earn the same revenue under any of those systems. The revenue will come from different sources in each case: in the ideal spot market, it will all come from energy sales, while in the ICAP or operating reserves systems, it will come in part from capacity/reserve payments, making up for lower energy revenues. The incentive for retirement or continued production by an existing plant will be the same in every case. The argument also applies to new entrants: their net revenues from capacity/reserve payments and energy sales will be the same under any of those systems under the assumptions made.

These conclusions might not follow in a market in which there is market power. Nevertheless, the conclusion that there is no bias in the competitive case indicates that an ICAP market does not necessarily prejudice retirement or entry decisions.

Entry of new LSEs is also an issue. There has been a complaint in PJM because a large fraction of the cost of power for new LSEs has consisted of capacity payments (PJM-Interconnection Market Monitoring Unit, 2000g), and that these payments have been predominantly made to owners of existing plants. As a result, new LSEs may view existing LSEs that still have capacity

[29] ISO-NE (2000b). Retirement decisions are affected by whether projected net revenues over the next few months or years cover the costs involved in keeping a unit available to generate, considering the costs involved in mothballing or decommissioning a unit. Long-term contracts or forward markets provide useful information for generator owners considering the option of retirement.

as having an unfair advantage. To the extent that the ICAP market inflates payments to generators, this can be true. However, since the analysis in §4 implies that an ICAP system does not generally increase the total cost of supply, we conclude that an ICAP system does not necessarily discriminate against entering LSEs who must buy ICAP credits. This is because, under the assumptions of §4, the average cost of energy from the spot market will be less under the ICAP system.

ICAP Has Only an Indirect Relationship to Reliability. By definition, ICAP regulation has the objective of assuring system adequacy, *i.e.*, to ensure that enough capacity is built and operable in the pool in the long run. ICAP payments in PJM are linked to historical performance (forced outage rates) of facilities. However, critics of ICAP point out that there is not a direct link between ICAP payments and availability during times when capacity is most needed for reliability purposes; instead, "iron in the ground" is rewarded.

In particular, ICAP payments are not related to the ability of resources to start-up and ramp-up quickly when needed (McCormick, 2000). Also, extra effort to increase availability during periods of greater need is not rewarded; in fact, if ICAP is coupled with an energy price cap, the incentive to quickly complete maintenance and schedule it away from peak periods may be weakened significantly. Finally, availability of generation capacity during peak periods can be compromised by the delisting of the resources from the pool, or as we will see, by exporting the energy to neighboring markets without delisting it from the system.

To the extent that measures can be taken to increase generator availability and responsiveness during just peak periods, the ICAP system does indeed seem to provide inefficient incentives. This is because ICAP, in effect, pays based on capacity availability throughout the year rather than when it is actually needed (Hanger *et al.*, 2000).[30] The analysis of §4 does not consider alternatives for increasing availability during peaks, and so its conclusions about the efficiency of ICAP systems need to be qualified. Quantitative analysis of the issue of investments and maintenance scheduling for improving availability at times of peak demand is desirable in order to assess the degree of inefficiency resulting from an ICAP system.

A fundamental fix to this problem is to reward availability and responsiveness during periods of short capacity in either of two ways:

1. replace the ICAP system with an operating reserves market that makes payments to reserves when most needed, considering start-up and ramp-rate limitations (Singh and Jacobs, 2000; see §3.2.2); or

2. replace the ICAP and spot market price cap with the Oren (2000)/Vázquez *et al.* (2001) mandatory call option system (§3.2.3).

Either system rewards greater availability and flexibility during times of capacity shortages. The first fix pays more for capacity that can serve as operating reserves at such times. The second fix rewards available generators during peak periods by allowing them to avoid having to pay an amount equal to the difference between the unrestricted spot price and the strike price.

[30] However, the PJM Market Monitoring Unit (PJM-Interconnection 2000d) has proposed changes that would provide incentives for delivery during peak.

Another criticism of the ICAP system relates to who is responsible for the maintenance of the system's reserves. Singh and Jacobs (2000) argue that compared to operating reserves-based systems, ICAP markets place all the responsibility for capacity reserves on LSEs, rather than diffusing the responsibility among all participants in the market, as an operating reserves market does. For example, if some customers pay real-time prices, an operating reserves system—with its attendant price spikes due to reserves shortages—will motivate customers to cut back just at those times when the system is at risk. This consumer incentive is removed if energy prices are capped, and only installed capacity is traded, with its cost buried in an uplift (Stoft, 2000b).[31]

A final criticism of ICAP under this category is that most ICAP payments are made to existing generators who would exist in any event (Strategic Energy, 2001). Payments to such "free riders" are viewed as being wasteful. It has therefore been proposed that payments only be made to new generators ("directed ICAP", *ibid.*). There are at least two arguments against this proposal. First, by fixing the total amount of ICAP credits required in the market and making all generators eligible for ICAP payments, there is more assurance made that sufficient capacity will be made available. This is because the directed ICAP scheme has no provision for requiring a certain amount of total capacity in the market. Second, ICAP payments to existing generators motivate them to undertake life extension and other improvements that will improve their capacity and availability.

3.3.2.3. Design-Dependent Criticisms of ICAP. The criticisms summarized in this section address particularities of the ICAP system design. Presumably, these criticisms could be addressed by modifying the design of the market The criticisms include: a poorly designed ICAP auction can magnify market power problems; the ability to delist ICAP capacity at short notice can harm system adequacy; and ICAP is an unnecessary complication if capacity cannot be sold on a firm basis outside the ISO. The first and third problems are associated with the New England ICAP market, while the second problem is a concern in PJM.

Market Power. Market power is the ability of market participants to affect prices. This ability can stem because the market is dominated by a few large firms, but it can also be a result of rules designed to govern the market. If the former is the cause, then market power is likely to occur in any market that those firms participate in, and there is no reason to believe that ICAP markets are inherently more prone to market power than, say, spot energy markets. Indeed, because the ICAP trading can be done on a long-term basis, entry and competition may be more effective than in hourly spot markets.[32] However, poorly written rules can make the capacity markets easier to game (Hirst *et al.* 1999).

In any market, market power is classically exercised by withholding output (or capacity, in the case of ICAP) whose cost is less than the market price or, equivalently, bidding above marginal

[31] On the other hand, the ICAP credit system provides an incentive for the LSEs to implement ALMs that lower their capacity requirements. To the extent that such ALMs can be used to reduce low-value loads during times of high energy prices, they can have much the same effect as real-time pricing programs. As we note elsewhere in this report, it is also possible to design real-time pricing programs in an ICAP system in which the cost of capacity could instead be recovered by price spikes during periods of shortages (Hobbs *et al.*, 2001b).

[32] The recent California experience apparently shows that market power is relatively easy to exercise in spot markets, especially if buyers are not allowed to engage in long-term contracts. To the extent that ICAP markets increase the amount of long-term contracting for supply, ICAP markets might actually dampen the exercise of market power.

cost. For instance, it has been claimed that holders of credits for active load management (ALM) have an incentive to withhold credits from the PJM ICAP market (McCormick, 2000). Such withholding in effect shifts the market supply curve to the left, raising prices. Further, if demand responds to price (the demand curve is not vertical), less total quantity will be supplied.

In networked markets, firms can also exercise market power by manipulating network constraints. For instance, local market power is possible if constraints prevent competing firms from importing power. In spot power markets, Kirchhoff's laws offer additional opportunities for manipulating prices, for instance by expanding output of strategically located plants so that lines linking a market to rival producers become congested or by altering outputs of generators in order to eliminate congestion payments to ISOs (*e.g.*, Berry *et al.*, 1999).   The PJM ICAP market is not subject to these types of manipulation because capacity is are generally defined on a pool-wide basis. (On the other hand, local ICAP markets, such as those in the NY pool, more accurately reflect the physical reality that extra capacity in a market is not worth much to a local area experiencing shortages because of transmission constraints.)

However, if transmission limits are severe within an ISO, it can make sense to define two or more geographically separate ICAP submarkets, as is the case in New York, discussed *infra*. As a result, effective market concentration can be very high in a submarket, increasing the potential for exercising market power. This potential can be mitigated by allowing external suppliers to participate in the capacity market up to the limits imposed by available transmission capacity into the area. PJM (PJM-Interconnection, 2000f) and NYISO (but not New England) allow for External ICAP. For such resources to discipline prices, there should be sufficient transmission capacity set aside. But in PJM, there are apparently too few transmission rights reserved for this purpose: 3500 MW are reserved as Capacity Benefit Margin for accommodating the 6500 MW that currently are imported (McCormick, 2000).   As a result of these transmission limitations, there may be too little generation capacity available to meet PJM loads during peak periods, and competition may be limited.

To analyze if there has been market power exercised in an ICAP market, transaction data should be available. In theory, the exercise of market power can be detected by submission of bids greater than marginal cost. However, "marginal cost" can be difficult to determine when that cost can be determined by foregone opportunities for sales in other markets rather than out-of-pocket expenditures associated with maintaining and operating a facility. For this reason, the "cost" of ICAP in PJM can have more to do with expectations regarding energy prices in other markets than with the out-of-pocket expense of keeping a plant available. Such expectations are difficult to estimate; as a result, it is difficult to determine whether ICAP bids deviate significantly from costs.   Nevertheless, the following types of bidding behavior are strongly suggestive of attempts to manipulate prices:

- erratic bids over time,
- very different bids for different increments of capacity,
- illiquidity—the unavailability of ICAP credits to willing purchasers (Goulet, 1999; McCormick, 2000), or
- failure to bid capacity when there is a region-wide surplus and ICAP prices are above the cost of keeping capacity available

Even if market bid and transaction data are absent, information on the number of participants and their shares of the market is useful as a rough indicator of the potential for market manipulation. (Indeed, the use of concentration indices lies at the Department of Justice approach to market power analysis.) In the remainder of this section, we review reports of market power in the PJM, New York, and New England capacity markets, emphasizing the particular features of those markets that may lend themselves to such manipulation.

Although New England has been the source of most complaints about ICAP market manipulation, there have also been such complaints in New York and PJM. We have not found individual transaction data from PJM, but there is an analysis by the PJM Market Monitoring Unit (PJM-Interconnection, 2000d) of the June 2000 ICAP price event. At that time, ICAP prices reached their highest levels since the ICAP market was established, and yet significant amounts of capacity were delisted (see Appendix V on a summary of the events on June 2000 in PJM ICAP market). However, the conclusion of the analysis was that PJM's ICAP market was not being gamed, and that the high bids can be explained by the opportunity costs associated with sales possibilities in other regions with higher (uncapped) prices (*i.e.*, ECAR). Nevertheless, market concentration in PJM continues to be viewed as a critical issue (McCormick, 2000), especially after the high ICAP prices in 2001.

In New York, the ICAP market design has yielded a high concentration of ICAP suppliers. In particular, the ICAP market is defined by the NYISO such that there is a special ICAP requirement for New York City. There are only four ICAP credit holders in that area, and prices are at their cap all the time (www.nyiso.com). The design creates this special area with high concentration of market and very few participants. Of course, as the Argentinean experience described *supra* shows, capacity that cannot contribute to system reliability because of transmission constraints is useless, so defining geographic submarkets for ICAP in this manner can make sense. (It is certainly nonsense to ignore transmission constraints altogether, as capacity in, say, Wisconsin cannot help with adequacy problems in New York City during peak periods when transmission is constraining.) The result of defining such markets, however, can be enhanced opportunities for gaming.[33]

Nevertheless, the New York City ICAP market is not believed to be gamed, due to a tight capacity supply in that region, and tight transmission, which limits external participation in the market. This conclusion follows from the fact that when supply is insufficient to meet a fixed demand, the equilibrium price is at the cap, even if the market is very unconcentrated.

New England, where the most complaints have been made about market power in the ICAP market, is the focus of the remainder of this section. Successful attempts to game the market have been reported there. The ISO-NE ICAP market was implemented on April 1, 1998. Because of dissatisfaction with the market's performance, on May 8, 2000, ISO-NE proposed the elimination of the ICAP market, effective June 1, 2000 in a filing with FERC (ISO New England Inc., 2000b). ISO-NE made this request following a comprehensive review of the ICAP market design and operation, including consideration of bidding activity that required modification by

[33] Revenues received by ICAP providers in the New York City ICAP market have been large in magnitude, and can be calculated from data posted in the NYISO web site (www.nyiso.com). This equals (8.75$/kW-month * 120.4 MW sold) = $1,053,500/Mo for the monthly auction cleared in October 2000 for the New York City region, where $8.75/kW-month is the price cap. For the summer 2000 New York City ICAP strip auction (May - October 2000) in the NYC region, 5,409 MW were sold at a price of $52.5/kW (=$8.75 x 6 months), or more than $25 million.

the ISO of certain ICAP market bids. Bids submitted for significant amounts of installed capability in January 2000 and subsequent months represented part of a pattern of behavior consistent with an intentional or systematic effort to raise the market clearing prices. ISO-NE's petition to drop the ICAP requirement was rejected by FERC, although permission is still pending to eliminate the monthly auction in which the most obvious gaming was taking place. FERC did, however, authorize a bid cap, as discussed below.

It was the poor design of the monthly "residual" ICAP auction in New England that made the gaming possible. In particular, the daily market had a fixed demand that was calculated by the ISO as equaling the sum of calculated shortfalls of ICAP for New England ISOs. These shortfalls were automatically inserted in the auction as, fixed quantity bids for capacity. This inflexible demand bidding arrangement invited gaming. Because there was no price elasticity and no bid cap, suppliers of ICAP were encouraged to submit very high bids in the hope that those bids might set the market clearing price. A better design would have made demand bidding voluntary, as in PJM, with penalties for having inadequate credits. Some details are provided below on the gaming that was observed in the New England market.

A particular trend in the bids for the ICAP market was observed during the June 1999 – January 2000 period. First, the highest bid in each month has increased from near $1,000 per megawatt to as high as $99,999/MW for significant blocks of ICAP (this was the highest bid allowed by the computerized bidding system). Second, for most of the period a large quantity was bid at $0 per megawatt and a small amount bid at high prices forming a "J" shaped supply curve. This type of behavior cannot be reasonably related to costs. This is because the only opportunity cost is the price in the New England bilateral market, as New England suppliers are required to sell their capacity in New England.[34] This opportunity cost should be low, since New England has ample capacity, as shown in Table 3.1. The supply curve shape changed somewhat in December and dramatically in January with significant quantities bid in the market at the highest price offered.

Table 3.1: Activity and Prices in the New England ICAP Market

	Aggregate Of Participants' Adjusted Monthly Peak Loads	Total Of Participants' Minimum System Capability	ISO-NE Capability Responsibility	Total Excess ICAP	Excess ICAP Sold To The Market (Pool Deficiency)	ICAP Market Clearing Price
Month / Year	(MW)	(MW)	(MW)	(MW)	(MW)	($/ MW)
January-00	21,088	27,065	24,772	6,547	4,254	$0
February-00	19,521	27,813	24,681	7,154	4,021	$0
March-00	17,532	28,376	24,778	6,401	2,798	$0
April-00	16,547	28,561	26,039	5,929	3,406	$3,248
August-00	21,986	27,466	26,322	4,287	3,144	$8,750
September-00	21,540	27,101	26,332	4,803	4,035	$8,750
October-00	18,088	28,876	27,452	4,945	3,521	$8,750

[34] In contrast, PJM suppliers are not compelled to sell their capacity in the PJM ICAP market if they prefer instead to sell it outside the region.

In January 2000, four New England market participants had bid authority for units with claimed capacity of 2000 MW or more.[35] In the absence of other obligations, any of these participants could have unilaterally set the clearing price for ICAP in some months. In other months there were only two effective bidders for the last megawatts of the Installed Capability requirement. The economic withholding of one such block in January, coupled with certain other bid behavior, led to mitigation actions by ISO-NE. The ISO retroactively set the ICAP clearing price for January to $0.00 on April 7, 2000 after the ISO had mitigated bids in the January 2000 wholesale ICAP market (ISO New England Inc., 2000a). The action was taken after the review of market conduct by the ISO under authority granted to it by market rules approved by FERC. In particular, Market Rule 17 states that the ISO can identify "generators whose pattern of behavior is consistent with an intentional or systematic effort to raise the market clearing price; and whose pattern of behavior, without mitigation, would actually cause a material increase in the market clearing price."

The ISO-NE believed that nothing protects the ICAP market from a recurrence of this behavior. Therefore, on September 18, 2000, ISO-NE's President and CEO, Philip Pellegrino, announced to the Participants Committee meeting that ISO-NE is intentionally holding back the Final ICAP Settlement for May-July 2000 until guidance from FERC was received. On December 15, 2000, FERC ruled on the ICAP deficiency charge to be used in ISO-NE market and set it to $8.75/kW-month (FERC, 2000b). The FERC order has an effective date of August 1, 2000. Therefore, ISO-NE settled the ICAP market as shown in the table below. Meanwhile, ISO-NE is moving towards the design and implementation of an improved operating reserves system (§3.2.2) to control short-term security and keep track of long-term adequacy.

<u>*Problems with Unit Delisting.*</u> ICAP's purpose is to ensure long-term adequacy. This implies that ICAP units are dedicated to being available to provide power to a particular market for some long period of time, such as a year. Different ISOs have different rules governing the nature of this commitment. NYISO has seasonal auctions (summer and winter), and bids are accepted for the periods, summer and winter periods. But some ISO resources in that ISO are exchanged as month- and day-ahead products, just as PJM conducts a daily and monthly market for this product. Bilateral exchanges can involve any terms mutually agreeable to the parties involved. Very short term trading raises the possibility that generators that are sold as ICAP during low load periods will be diverted to other markets ("delisted") during times when spot prices are higher elsewhere.

Such delisting is seen as a threat to reliability, and is arguably the most severe criticism of the PJM ICAP system. In PJM, ICAP owners may delist with two-day advance notice. If units are delisted, the ISO is not able to recall them, so in case of emergency, the pool resources are diminished. As a result, LSEs may pay most of the year for ICAP capacity that is not there when truly needed; as a result, PJM or the LSEs may have to buy very costly spot power from outside PJM in order to meet their obligations. In a sense, this is paying twice for reliability: once to ICAP providers, and a second time to spot power providers during times of price spikes. A further risk is that capacity owners might use the delisting mechanism manipulate the market. Conditions under which strategic market manipulation is more likely to happen should be explored in further research.

---

[35] See www.iso_ne.com/power_system/documents/net_claimed_capability_report.

Delisting is most likely to be a problem when nearby markets have different capacity market designs (Goulet, 1999; Stoft, 2000a,b). Power pools and markets are not isolated, independent physical entities, in general. The eastern US is linked through high voltage transmission lines, so the injection of energy by a generator in, for example, the ECAR area affects energy distribution in PJM. There are no jurisdictional or state boundaries for the flow of energy. But different ISOs apply different market rules. As it is possible for a generator to offer energy for sale in markets outside its control area, different markets interact. This interaction creates situations that can affect a market's capacity adequacy.

Stoft (2000a) explains how this interaction allows price spikes to be imported from one market to other, even when there are rules designed to prevent this from happening, and how this interaction can deplete markets of generation capacity when they need it the most. Stoft refers to two market failures that can contribute to this effect: market rules (regulator intervention) and market power.

The market rules problem arises when one ISO has an ICAP system that allows delisting along with an energy price cap (as in PJM) and a nearby market has a higher or no energy price cap. As a result, ICAP suppliers in the former market have an incentive to delist when prices are higher in the latter market. This can happens as follows. The highest marginal generating cost in PJM from April 1998 to April 1999 was about 155$/MWh, while in the following year it was 130$/MWh. (The highest bid, of course, was much higher, since bids did not have to be cost-based.) If generators who have committed capacity resources to the PJM ICAP market observe peak prices in excess of the PJM cap of 1000$/MWh in a neighboring market while prices are lower in PJM, they have a motivation to sell outside PJM to take advantage of the price difference. They can do this delisting the generating unit (with a required but short warning time).

The recall process works as follows. When PJM declares a Maximum Generation Emergency (see Appendix III), it can recall ICAP generators who are selling outside PJM after accepting every energy-bid in the PJM day-ahead market. If in that case the owners of the exporting generators chose not to make them available to provide power to PJM, penalties are imposed. Because these penalties have been relatively low (at most, double the levelized daily cost of a combustion turbine, see PJM Market Monitoring Unit (2000a) together with an upward adjustment in the forced outage rate that PJM assigns to that unit), it has become profitable for such generators to "delist" in this manner and make themselves unavailable to PJM pool. But because of some wrinkles in the PJM energy market rules, there is some disagreement as to whether this effect has been observed. Stoft (2000a) argues that it has not been clearly established, while the PJM Market Monitoring Unit (PJM-Interconnection, 2000g) appears to disagree.[36]

---

[36]This is in part because bidders in PJM are no longer constrained to bid their marginal cost. As a result, bidders can raise their bids to reflect the opportunity costs of sales in other markets. Indeed, because of a quirk in PJM pricing rules, the PJM energy price can be artificially raised above market clearing levels in some situations. In particular, as just noted, PJM rules say that resources bid within PJM must be used before recalling generators selling from PJM to outside the region, buying spot imports, or taking other emergency actions. The price paid to bidders is then the highest within-PJM bid (as high as $1000/MWh), even if, for instance, price of spot imports is well below that level (say, $300/MWh in ECAR). As Stoft (personal communication) points out, this provides an incentive for PJM LSEs with extra generation to commit generation outside the region in order to induce the shortages that would invoke this rule. An additional, even stranger result of this rule is that bidding an additional generator into PJM can

The PJM Market Monitoring Unit (MMU) (PJM-Interconnection, 2000g, p. 52) concluded that operations should be modified to eliminate participant's incentive to export energy. In their State of the Market during 1999 report, Point 3.4 proposes short-run rule changes in response to observed price spikes:

> "PJM should consider devising a system with capacity resource deficiency penalties that vary over time. Rather than a fixed penalty per day throughout the year, the penalty would be higher when the value of capacity resources is higher."

However, any proposed change in the penalty will be ad hoc to some extent, as there is no accepted basis for determining the "right" penalty. The basic problem is the fundamental incompatibility of different types of remedies for the failure of energy markets to provide sufficient capacity, complicated by price caps and other market rules. One solution is to have uniformity among regions in their treatment of capacity (which could occur, for instance, by having PJM abandon energy price caps and ICAP); however, such uniformity is unlikely given FERC's tolerant approach to Regional Transmission Organization design. Another solution is to develop a penalty for delisting that reflects the true underlying costs of the system.[37] For instance, penalties in any given day could be based on "liquidated damages" which would reflect the actual costs that PJM incurs dealing with emergency conditions. The Oren (2000)/Vázquez *et al.* (2001) mandated call-option approach (§3.2.3) can be viewed as such a penalty system—the penalty for delisting is to make the holder of the ICAP right whole. But for the moment, PJM proposed a modified penalty scheme in the filing with FERC (*Proposed RAA Schedule 11 § B.3*, PJM-Interconnection, 2001e) that calculates fines assessed for a interval of time (each year divided in 3 periods: a 4-month period starting June 1st, a 3-month period from October 1st, and a 5-month period starting Jan 1st), rather than on a daily basis. This scheme takes into account the maximum capacity deficiency that the LSE had in one day during that interval multiplied by the duration of the interval and the Deficiency Rate (expressed in dollars per MW Unforced Capacity). This is not a direct application of the above MMU proposal, but does increase the penalty for summer and winter months.

ICAP is an Unnecessary Commodity with Zero Opportunity Cost. The last design-particular criticism is directed at the ISO-NE ICAP design. Because the market there in effect requires all New England capacity to be ICAP and prohibits delisting, it is argued that (at least in the short run) there is no value to an ICAP resource because it has no opportunity cost outside New

---

actually *increase* the price, even if all other bids are the same (Stoft, ibid.). (In normal markets, increasing the supply in this manner cannot increase prices, and may decrease them.) If there is a shortage, and the last bid unit in PJM costs $600, then all bid units receive $600; but if someone had in addition submitted a very small plant with a bid of $950, then that is the price all units would receive.

However, the PJM Market Monitoring Unit (PJM-Interconnection, 2000g) shows compelling evidence of capacity withdrawal as a result of external prices well in excess of PJM prices during July 29-31 and Aug. 14-15, 1999. On one of those days, for instance, the external energy price reached $2000/MWh compared to PJM's $400/MWh. As a result, capacity imported into PJM disappeared, and less capacity was made available by entities within PJM. That same report predicted that such price differences would lead to run-ups in PJM ICAP prices—as indeed they apparently did in June 2000.

[37] The Brattle Group (2000) analyzed the incentives for producers in the NYISO to migrate their capacity from the pool when a neighboring market has very high prices, in the context of reducing their Obligation Procurement Period from six months to one month. Under energy caps, the situation changes as observed by Felder (2000) in a Memorandum to the NYISO.

England. This is the argument of made in an affidavit filed by Peter Cramton together with the ISO-NE petition to eliminate its ICAP market before FERC (ISO New England Inc., 2000b). Cramton maintains that the New England ICAP clearing price should be the marginal cost of having one more unit of capacity available, which in the short-run is very low.

ICAP is defined differently in New England and PJM. Most importantly, in PJM an ICAP resource is recallable during capacity shortages. However, PJM generators have the option to (1) ignore the ICAP recall and pay a penalty or (2) delist from ICAP and not be subject to recall. In contrast, in New England, all generators could be called in case of insufficient generation available, even if they did not submit bids to the ICAP market or commit the capacity in the bilateral ICAP market. So there is no opportunity cost in New England for the generators, as all resources are, in effect, ICAP, and they can only trade them in the bilateral or ISO-NE ICAP market. So Cramton's argument that there is no opportunity cost applies only to this specific market design.

In the long run, however, if the ICAP requirement is binding (no excess capacity), then the price of capacity should rise to the price cap, and provide an incentive for expansion. However, New England does not now have a capacity shortage, so the ICAP payments should, in theory, be zero and provide no incentive for capacity additions. There are more than 30 GW of capacity proposed to be installed in the pool, much more than what it is needed.

4. A STYLIZED COMPARISON OF CAPACITY MARKET ALTERNATIVES IN PJM

In the previous sections of this report, we have made various assertions about the relative economic efficiency of ICAP, spot market, capacity payment, and operating reserves systems, and their effect on system adequacy and incentives for adding or retiring capacity. The basic assertion is that under pure competition, it is possible to design any of these types of capacity market systems so that a predetermined level of adequacy is achieved at least cost. None of the systems necessarily favors or discourages retirement or additions of capacity more than the other systems. The purpose of this section is to justify those assertions with a simple probabilistic model of an energy and capacity market involving two types of generation facilities.

Similar types of analyses have been undertaken before. In an attempt to clarify the mechanisms of and differences among alternative market designs, Stoft (2000b) uses a deterministic market model to represent the ideal pure energy market, an ICAP-type system, and an operating reserves type system. Inelastic consumer demand is assumed, as is perfect competition (*i.e.*, no generator believes it can alter the price). The analysis in this section is inspired by his work, and extends it in several ways. The main extension is to consider all systems under the same set of assumptions concerning variations in demand and randomness of generator outages. This stochastic analysis allows us to explicitly relate each system to the resulting LOLP. We also consider the question of how the mix of base- and peak-load facilities could be affected by different capacity market designs; Stoft (2000b) considers only peaking facilities. Finally, our market designs differ in some ways from his in order to make the comparisons simpler, and we also consider price-based designs.

The basic conclusion of our analysis, however, is broadly similar to his: in the long run, very different capacity market designs can yield the same level of system reliability (expressed as LOLP) and costs. Whether price-based or quantity-based, whether focused on installed capacity or operating reserves, it is not possible to say that one of the alternatives is inherently superior to the other on this basis. Our conclusions also go beyond Stoft (2000b) in that we find that the mix of base- and peak-load capacity can be the same; there are no systematic distortions that necessarily result from one system or another. Consistent with Stoft (2000b), however, we do show that the volatility of prices and prevalence of price spikes do differ among the systems; however, total expenditures by consumers are the same.

In the following subsections, we outline some basic assumptions of the stylized system (§4.1), summarize the model (§4.2), describe the demand and generator data (§4.3), and present the results (§4.4). The results sections compare the generation capacity amounts and mixes resulting from spot market alone (§4.4.1) and different capacity market proposals (§4.4.2). The resulting price distributions and volatility are discussed §4.4.3 and the effects on retirement decisions for existing capacity, in §4.4.4. The chapter concludes with a sensitivity analysis that illustrates how prices and equilibrium capacity mixes are affected if price spikes occur when demand approaches within a few hundred MW of available generation capacity, rather than only when an actual capacity deficit occurs (§4.4.5).

4.1. Market and Model Assumptions

We assume that in the long run, there are two types of generating units that can be used to meet PJM energy demands: peaking units (low capital cost, high operating cost) and baseload units (high capital cost, low operating cost). Each type of unit is also characterized by a fixed capacity size, and a forced outage rate. We disregard transmission constraints, so we assume there is a single market for energy (and, under some market designs, operating reserves and installed capacity). The load duration curve describing the annual distribution of demand is adequately approximated using a mix of two normal distributions. We assume that for purposes of reliability and generation mix calculations, we can disregard planned generation outages.

If the probability of forced outage of different generators is independent, then the probability distribution of available capacity of a given type follows a binomial distribution. Furthermore, for a large system such as PJM, a normal distribution can be used to approximate this distribution. If n is the number of generators of a given type, FOR their probability of being unavailable if called upon, and CAP is their capacity, then the normal approximation has mean $n(1\text{-}FOR)CAP$ and variance $n(1\text{-}FOR)(FOR)CAP^2$. (This variance formula is also used in the PJM GUS Manual to describe the reliability of generators (PJM, 2001j)). The total available capacity is the sum of the (normally distributed) available baseload capacity and the (normally distributed) available peaking capacity. Such a sum is also normally distributed; assuming independence, the total available capacity has mean $n_1(1\text{-}FOR_1)CAP_1 + n_2(1\text{-}FOR_2)CAP_2$ and variance $n_1(1\text{-}FOR_1)(FOR_1)CAP_1^2 + n_2(1\text{-}FOR_2)(FOR_2)CAP_2^2$, where the subscript "1" refers to baseload plants and "2" refers to peaking plants.

If the available capacity is less than the demand, then there is unserved energy. There is a true value of unserved energy VUE^* (associated with the loads that go unserved), and a value that is expressed in the (imperfect) energy market VUE^o. VUE^o will be based on the WTP of LSEs to avoid load shedding, and is also affected by price caps. When there is such a shortage, the energy price is assumed to spike to the cap or VUE^o, whichever is lower.[38] We refer to such periods of price spikes as "Price Regime 1."

In all market designs except the operating reserves-based market, the energy price during times of capacity surplus equals the marginal cost of the most expensive unit serving the load (MC_1 if no peakers are being used or MC_2 if peakers are dispatched). Price Regime 3 is defined as occurring if peakers are the marginal unit, while Price Regime 4 obtains if baseload units are marginal. Price Regime 2, when energy prices are affected by the operating reserves market, only arises in the operating reserves proposal, which we explain later.[39]

We now describe the four basic alternative market designs.

[38] A more sophisticated representation (*e.g.*, Hirst and Hadley, 1999) would recognize that VUE would increase as the magnitude of the shortage worsens. Inclusion of this complication would not alter the basic insights obtained in this analysis. The increase in prices during times of shortage does not assume that bidders can predict the shortage. Instead, it is assumed that there are generators who submit high bids (equal to or exceeding the cap or VUE^o, whichever is less), and that there demand-side bid schedules are submitted whose highest prices are VUE^o.

[39] Oren (2000) also presents an example in which there are two types of generators, and price either equals the MC of the marginal generator or a VUE.

Pure Energy Market. In a pure energy market in which there are no payments for operating reserves, a generator receives the market price for energy (as described above) if it produces power and no other revenues. We are unaware of the existence any pure energy markets, but it is a useful benchmark for comparison. (For instance, PJM actually makes some payments for reserves, but only to cover otherwise uncompensated costs.)

Price-Based System. A price-based system of capacity payments assumes that the ISO pays P_{PB} \$/MW/year for "unforced" capacity. "Unforced" here is used in the sense that PJM uses it: as (1-FOR) times the capacity of the unit. Thus in addition to energy revenues, a generator obtains P_{PB}(1-FOR)CAP in capacity payments. In this system, and in the ICAP system, we also assume that PJM makes no payments for operating reserves (as in the current system). However, the generator does also receive the market price for energy, as defined earlier, if it produces power.

Quantity-Based System: Installed Capacity. An ICAP market is represented by a system in which the ISO requires that the total unforced capacity in the market be at least equal to $(1+RM_I)D_{max}$, where RM_I is the (fractional) unforced reserve margin and D_{max} is the peak demand.[40] We assume that ICAP credits are freely traded at price P_{ICAP} \$/MW/yr. It can be shown that a generating company that is maximizing expected profits will act as if it receives an annual payment of P_{ICAP}(1-FOR)CAP for its capacity, whether or not it is directly responsible for serving load (*i.e.*, whether or not it is an LSE).

Quantity-Based System: Operating Reserve Market. The ISO here creates a market for reserves. It sets a target reserve level of at least $RM_o\ D_{max}$ MW (where RM_o is the target expressed as a fraction of peak demand D_{max}). We assume as a rough approximation that all available peaking capacity is available for reserves if it is not generating energy, and that the marginal cost of providing reserves is essentially zero for existing idle capacity. Start-up and spinning costs are therefore ignored. As a result, if available capacity exceeds demand by more than $RM_o\ D_{max}$, the market clearing price of reserves should be zero, since they are in surplus and incur no incremental costs. The key assumption concerns what happens if there is insufficient capacity for reserves. First, we assume that, as in ISO-NE and California, there is a ceiling P_{OR} on the price that the ISO will pay for operating reserves. Further, we assume that bidding will raise the price of reserves to that point if there is less spare capacity available than the reserve requirement.[41] Later, we also consider a more sophisticated operating reserves proposal in which an ISO would pay a price for capacity or reserves that would increase with the degree of shortage (Stoft, 2000b). The New England ISO has proposed an operating reserves market of the latter sort. The simulation defines a linear pricing function for operating reserves that is zero when the reserves are exactly the target value described above, and ramps up to P_{ORmax} when the reserves are zero. This function reflects the increasing value that operating reserves have as capacity becomes scarce.

---

[40] RM_I is equivalent to PJM's FPR (forecast pool requirement); see Appendix II. Here, we assume that ICAP is an annual rather than daily obligation that cannot be withdrawn on short notice, consistent with the PJM Market Monitoring Unit (PJM-Interconnection, 2000g) recommendation.

[41] Competitive markets for operating reserves in which demand is perfectly inelastic, as it is here, will work in that manner. Stoft (2000b) uses a more sophisticated assumption in which the ISO has a demand curve for reserves that increases the price for reserves as the shortfall in reserves grows. Oren (2000) also discusses this approach. Our model of the operating reserves system, however, captures the essence of how this type of market operates, and is sufficient to illustrate how it affects capacity investment decisions.

As a result, there are four distinct price regimes in the operating reserves system:

1. Regime 1. Available capacity is insufficient to meet demand, so the energy price goes to VUE^o and no reserves are supplied. (We assume VUE^o is well in excess of P_{OR}, so no generator would prefer to supply reserves instead of energy in this case.)
2. Regime 2. Capacity can meet just the energy requirement, but is inadequate to meet the operating reserve requirement. In this case, the reserve price hits P_{OR}. Because generators are free to choose between the reserves and energy market, this effectively arbitrages between the two markets, so that the net revenue earned by a peaking unit must be the same in the two markets. As a result, the energy price must be $P_{OR}+MC_2$—that is, just enough higher to cover the cost of generation. (This assumes that peakers are the marginal sources of energy and reserves.)
3. Regime 3. Capacity is ample to meet both energy demand and reserve requirements, and peakers are the marginal generating unit. The energy price is then MC_2, and the reserves price is zero.
4. Regime 4. The baseload units alone can supply all the energy demand, so the energy price is MC_1.

In general, regimes 3 and 4 will constitute over 99% of the hours of the year, if there has been adequate investment in new capacity. Regime 2 will occur for most of the remaining hours, assuming that out-and-out capacity shortages are relatively rare.

4.2. Modeling Methodology

The basic approach is to consider a firm that is deciding whether to add capacity of a given generation type. If the firm assumes that it cannot affect the probability distribution or level of prices, and if it is risk neutral, it will construct such a unit only if its expected revenues exceed its expected cost. Thus, in equilibrium, an increment of capacity of each type will just break even; and since all units are identical and face the same prices, all units will break even. This condition allows us to write two equations representing these breakeven conditions:

For peaking units:

Expected revenue of the unit = Expected capital and operating cost (1)

For baseload units:

Expected revenue of the unit = Expected capital and operating cost (2)

Expected revenue includes energy revenues; revenues from installed capacity (either from payments P_{PB} or revenue P_{ICAP} from sales of ICAP credits); and/or revenues from selling reserves (during price regime 2), depending on the market design. Expected costs include fixed expenses (the annualized cost of construction plus fixed O&M costs) and variable operating costs (both fuel and nonfuel).

For the energy, price-based, and operating reserve systems, these equations can be written as nonlinear functions of just two variables: n_1 and n_2, the numbers of each type of generating unit. Because we use normal approximations for available capacity, the n_i are treated as continuous rather than integer variables. The nonlinear functions we use result from the normal approximations and the mix of normal approximation of the load duration curve. Using

EXCEL™ Solver's capabilities to find roots of equations, we solve this two equation-two variable system to find the number of generators of each type that result in the breakeven conditions being satisfied. These solutions will depend on the assumed generator characteristics, the distribution of demand, the value of unserved demand, and the assumed payments for installed capacity or operating reserves, if applicable.

However, for the ICAP market, the solution process is more complicated, since P_{ICAP} is not predetermined, but is instead an endogenously determined price that clears the capacity market. As a result, a third condition is needed to solve for this third unknown variable. This third condition expresses the market clearing condition:

$$\begin{aligned} &\textit{If } P_{ICAP} > 0, \textit{ then } n_1(1-FOR_1)CAP_1 + n_2(1-FOR_2)CAP_2 = (1+RM_I)D_{max} \\ &\textit{If } P_{ICAP} = 0, \textit{ then } n_1(1-FOR_1)CAP_1 + n_2(1-FOR_2)CAP_2 \geq (1+RM_I)D_{max} \qquad (3) \end{aligned}$$

That is, the price of ICAP credits can only be positive if there is not a surplus of unforced capacity relative to the amount required by the ISO.[42] This is technically not an equation; instead, it is called a "complementarity condition"; but like a regular equation, it can be used to solve for unknowns. Thus, with two breakeven conditions and the market clearing condition, we are able to solve for the amount of each type of capacity along with the P_{ICAP}.

It turns out that if the spot market price VUE^o during capacity shortages equals the true value of unserved energy VUE*, then energy revenues alone (pure spot market) result in the capacity mix that maximizes social net benefits. That is, the amount of peaking and baseload capacity that is provided by the competitive market is the amount that minimizes the sum of customer outage costs, capacity expenses, and variable generation costs.

4.3. Demand and Generator Data Assumptions

Table 4.1 shows the assumed values of the generator parameters.

- The forced outage rates are broadly consistent with average PJM experience
- Generator sizes are slightly above those typical for new coal and combustion turbines, in part to reflect the possibility of generating units at the same plant being forced out because of common mode failures.
- Marginal costs for base load facilities reflect output-weighted cost for plants of those types in PJM, while the peaker MC is more typical of new facilities.
- Fixed costs F1 and F2 are based on EPRI Technical Assessment Guide data along with PJM assumptions about the cost of new turbines.

---

[42] As noted in §3.3.1, ICAP prices in PJM have been observed to be positive even when there is surplus capacity. This is at least in part due to the O&M costs associated with keeping capacity available, the opportunity cost of capacity which could otherwise be free to produce energy for other regions, and perhaps market power. We disregard these considerations here.

Table 4.1: Generator Characteristics

FOR_1 =	0.12	[ ]	FOR_2 =	0.05	[ ]
CAP_1 =	1200	[MW]	CAP_2 =	150	[MW]
MC_1 =	25	[$/MWh]	MC_2 =	45	[$/MWh]
F_1 =	140,000	[$/MW/yr]	F_2 =	63,000	[$/MW/yr]

We used 1996 PJM hourly demand data to describe the PJM load duration curve, the latest available to us at the time the analysis was undertaken. (Use of more recent data would not change the conclusions of this analysis.) The mean demand was 27,710 MW, with a standard deviation of 5762 MW. The peak demand was 48,524 MW. However, this distribution was highly nonnormal, with a much thicker upper "tail" than a normal distribution. That is, there were many more hours of high loads than a normal distribution with that mean and standard deviation. However, when we tested a "mixture of normals" approximation (Gross *et al.*, 1988; Mira and Sanchez, 2000), an excellent fit with just two constituent normals resulted. (A mixture of normals distribution results if a given variable has probability p_i of being drawn from normal distribution $N(\mu_i, \sigma_i)$; in our case, i = 1,2.) The following table shows the parameters of the approximation:

Table 4.2: Parameters of Normal Distributions Used in Load Distribution Approximation

	i = 1	i = 2
Probability p_i	0.9626	0.0374
Mean μ_I	27109	42916
Standard deviation σ_i	4918	2399

The parameters were chosen so that the fitted distribution had the same overall mean and standard deviation of the PJM loads, and also passed through certain points on the load duration curve in the peak area. Figure 4.1 shows that the fit to the actual 1996 load distribution is excellent, especially in the peak.

For the reliability target, we assume that the 1 day in 10 year LOLP standard of the PJM Reserves Requirement Manual is desired (PJM-Interconnection, 1997d), which we interpret as 24 hours out of 87,600 hours.

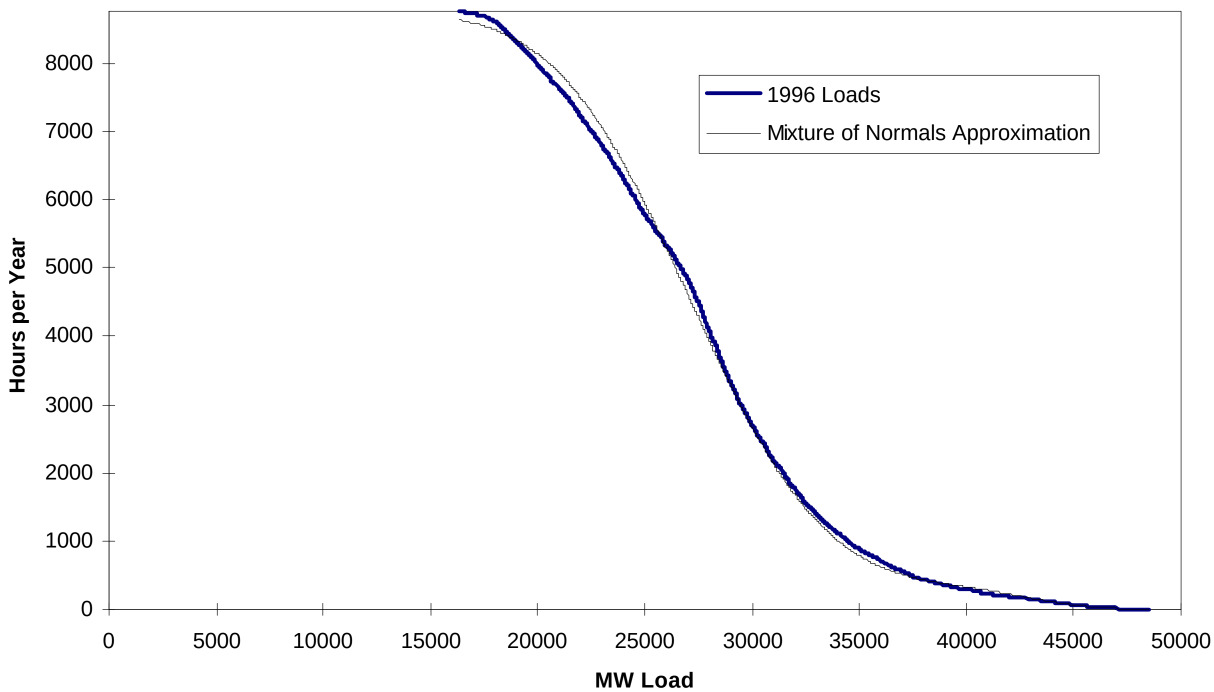


Figure 4.1: Comparison of 1996 Load Distribution and Its Approximation

4.4. Results

Several categories of results are reviewed below, including:

- the effect of the value of unserved energy/energy price cap upon the spot market results (§4.4.1);
- capacity mix and cost results for the quantity- and price-based systems (§4.4.2);
- the effects of alternative capacity market design upon price volatility (§4.4.3);
- the effects of alternative market designs upon the retirement decisions for existing plants (§4.4.4); and
- the sensitivity of price spikes to market behavior that occurs when demand approaches within a few hundred MW available generation capacity, rather than only when an actual capacity deficit occurs (§4.4.5).

4.4.1. Spot Market Alone

In a power market in which energy is the only source of revenues for generators, the spot price during capacity shortages VUE^o is critical in determining the amount of capacity and reliability. If VUE^o equals about $27,500/MWh, an admittedly very high value, this would yield a LOLP of

1 day in 10 years for this system, as the below table shows.[43] Over 56,000 MW of capacity would be provided to meat the peak demand of under 50,000 MW; the unforced reserve margin would be a relatively thin 4.2%.[44] If the true value of unserved energy VUE* is indeed $27,563, then this would be the optimal capacity mix.

Table 4.3: Pure Spot Market Results

VUE* [$/MWh]	Peaking Capacity [MW]	Baseload Capacity [MW]	Unforced Reserve [MW]	Unforced Reserve Margin [%]	LOLP [1/10yr]	P(Regime 1): Capacity Shortage	P(Regime 3): P = MC_2	P(Regime 4): P = MC_1
27,563	24,862	30,636	2,054	4.20%	1.00	0.03%	52.93%	47.05%
10,000	20,006	30,636	825	1.70%	2.80	0.08%	52.88%	47.05%
1000	19,508	30,636	(3,032)	-6.25%	28.90	0.79%	52.16%	47.05%
500	17,463	30,636	-4975	-10.25%	60.70	1.66%	51.29%	47.05%

The table shows what transpires if VUE^o instead takes on lower values: unforced reserve margins fall (ultimately to negative levels), and reliability deteriorates. Price spikes associated with capacity shortages occur more often—for example, for 0.8% of all hours if VUE^o = $1000/MWh.[45] However, the amount of baseload capacity is unaffected; only peaking facilities

---

[43] This value is consistent with the assumption that the annualized cost of a combustion turbine is $63,000 per MW (or $66,000 per "unforced" MW, at a FOR of 5%). If LOLP is 1 day in 10 years, then 2.4 hours of capacity shortage should be experienced per year. The turbine would have to make enough money in those hours to pay off its fixed costs; thus, the required price spike is on the same order of one-half the entire annual capital cost. Of course, if the reliability standard was relaxed to, say, allowing 24 hours of shortage per year, then the required price spike would be roughly one-tenth that level, which is more consistent with recent experience in Midwest and eastern US markets.

[44] This unforced margin is less than the 7.8% that PJM calculates is necessary to maintain its desired reliability level. Possible reasons for this include the following:

- Our disregarding of planned outages during the fall and spring month, which as shown in Appendix IV amounts to as much as 20% of the system capacity. Such outages can decrease system reliability, as was illustrated by the price spike of May 2000;
- We assume that the load duration curve and its peak is known. In reality, a distribution with a higher variance and peak should be used to reflect uncertainties in weather, including the possibility of more severe summer temperatures than experienced in 1996;
- We disregard the effect of transmission bottlenecks, which in general increase the reserves required to achieve a given LOLP; and
- We do not consider the need for additional capacity for regulation and spinning reserve; since system operators will shed load when those go below critical levels, it is necessary to add perhaps 5% to system capacity to accommodate those needs (Hirst and Hadley, 1999).

This discrepancy between the unforced reserves PJM and we calculate, however, does not affect our basic conclusions about the long run efficiency of alternative capacity market designs.

[45] Compare this to the 71 hours during the March 1999-February 2000 period (0.8%) that Hanger *et al.* (2000) report that PJM prices exceeded $400/MWh. If outages occur this often with prices in the $500-$1000/MWh range, combustion turbines can earn enough to cover their annualized capital cost. (*E.g.*, $750/MWh*50 hours/yr =$37,500/MW/yr, the same order of magnitude as the cost of a turbine.)

are impacted. The net social cost of each solution is not shown, but if the true VUE* is $27,563, the other solutions result in significant increases in social cost—the result of the market failure that VUE^o does not equal VUE*.

4.4.2. The Performance of Price-Based and Quantity-Based Systems

If the target reliability of the system is LOLP = 1 day in ten years, this can be accomplished using any of the three alternative capacity market designs: price-based, ICAP, and operating reserve requirements. Assuming that energy prices are capped at \$1000/MWh (=VUE^o), the following table shows versions of each market design that accomplish that goal[46].

Table 4.4: Comparison of Alternative Market Designs that Achieve LOLP = 1 [day/10 yrs][a]

Capacity Market Design	Peaker Capacity [MW]	Baseload Capacity [MW]	Unforced Reserve Margin [%]	LOLP [1/10yr]	P(Regime 1): Loss of Load	P(Regime 2): Capacity Shortage	P(Regime 3): $P = MC_2$	P(Regime 4): $P = MC_1$
ICAP (P_{ICAP} = \$64,000/MW/yr)	24,855	30,636	4.20%	1.00	0.03%	n/a	52.93%	47.05%
Price-Based (P_{PB} = \$64,000/MW/yr)	24,855	30,636	4.20%	1.00	0.03%	n/a	52.93%	47.05%
7.5% Operating Reserves (VUE^o = \$1000/MWh; P_{OR} = \$2070/MWh)	24,867	30,635	4.24%	1.00	0.03%	0.35%	52.57%	47.05%
7.5% Operating Reserves (VUE^o = \$2030/MWh; P_{OR} = \$1985/MWh)	24,863	30,635	4.23%	1.00	0.03%	0.35%	52.57%	47.05%
7.5% Operating Reserves Variably Priced (VUE^o = \$5150/MWh; $P_{OR\ max}$= \$5005/MWh)	24,871	30,635	4.25%	1.00	0.03%	0.35%	52.58%	47.05%

a. All solutions assume energy price cap (VUE^o) = \$1000/MWh, except last Operating Reserve case

The first row of the table reveals that the ICAP system can reach the LOLP goal by directly imposing an ICAP requirement of 4.2% (unforced reserve margin). This would yield an ICAP credit price of \$64,000/year (\$175/MW/day). The second row shows that this same amount, paid

---

There are at least two reasons why price spikes actually occur for more hours than indicated here. First, this analysis does not consider the fact that prices will actually start spiking upwards when capacity is less than regulation plus spinning reserves requirements, which will occur for more hours during the year. Second, what we call "loss of load events" will, in most cases, be dealt with by curtailing interruptible loads and invoking load controls without actually having to resort to involuntary outages. These events will also result in price spikes as LSEs seek to avoid such curtailments, and will occur much more often than 1 day in 10 years.

[46] We disregard the complications arising from multisettlement systems in which uncertainty in the forward market could lead to run-ups in forward and bilateral prices even when a shortage is not actually realized.

in a price-based capacity payment system, would elicit the same unforced reserve margin. The third row shows that if an operating reserve margin of 7.50% was enforced, and the ISO paid $2070/MWh for such reserves if they are short supply, then the same LOLP would result. Such a payment for reserves, however, is unlikely to be made by the ISO if at the same time it imposes a cap of $1000/MWh for energy. If the operating reserve payment were lowered to $1000/MWh, the LOLP would rise to 3 days/10 years. The fourth row of the table shows an alternative operating reserves solution: raise the energy price cap to $1985/MWh, and also pay $2030/MWh for reserves. This difference ($45/MWh) between the energy price and the reserves prices is necessary in order to make the turbine indifferent to generating and providing reserves. This achieves the LOLP target. Finally, the last row shows that a market with variably priced operating reserves with a 7.5% reserve margin can also be designed to achieve the same LOLP, which then yields the same generation mix as the above systems.

Note that the table reveals that each of the solutions achieves the targeted 1 day in 10 years, and their generation mixes are almost exactly the same. Furthermore, this mix is the same as the ideal spot market in which $VUE^o = VUE^* =$ $27,563/MWh. Thus, in theory, each of the systems can achieve the reliability goal at least cost, confirming Jaffe and Felder's (1996) argument that price-based and quantity-based systems are theoretically equivalent. The exception is if the price of reserves is capped at such a low level that there is no incentive for producers to add that amount of capacity.

It is of interest to compare the sources of each type of generator's revenue. If we define gross margin on sales as the excess of revenue over cost, each generator must earn enough gross margin to cover its fixed cost. Table 4.5 shows the sources of gross margin in each proposal for peaking facilities, while Table 4.6 shows the same data for base load plants.

Table 4.5: Gross Margin (in $/MW of capacity/year), Peaking Plant

Market Design	Regime 1 (Energy Shortage)	Regime 2 (Operating Reserve Shortage)	Regime 3 (Peaker is Marginal Unit)	Regime 4 (Baseload Plant is Marginal Unit)	Capacity Payment (P_{ICAP}, P_{PB})	Total
Energy Spot Market Only (Table 4.3, Row 1)	63,000	n/a	0	0	n/a	63,000
ICAP (Table 4.4)	2170	n/a	0	0	60,830	63,000
Price-Based (Table 4.4)	2200	n/a	0	0	60,800	63,000
Operating Reserves (Table 4.4, Row 4)	4544	58,456	0	0	n/a	63,000

Table 4.6: Gross Margin (in $/MW of capacity/year), Baseload Plant

Market Design	Regime 1 (Energy Shortage)	Regime 2 (Operating Reserve Shortage)	Regime 3 (Peaker is Marginal Unit)	Regime 4 (Baseload Plant is Marginal Unit)	Capacity Payment (P_{ICAP}, P_{PB})	Total
Energy Spot Market Only (Table 4.3, Row 1)	58,400	n/a	81,600	0	n/a	140,000
ICAP (Table 4.4)	2,053	n/a	81,600	0	56,347	140,000
Price-Based (Table 4.4)	2,080	n/a	81,600	0	56,320	140,000
Operating Reserves (Table 4.4, Row 4)	4,250	54,695	81,055	0	n/a	140,000

Several observations can be made from these tables. First, in the spot market system, the very high spot price during Regime 1 not only accounts for all the peaker's gross margin (as must be the case), but also a very large share of the baseload unit's.[47] In the operating reserves system, the same is true of Regime 2, showing that generators will be depending heavily on high energy prices a few hours per year to cover much or all of their capacity cost. In contrast, in the price-based or installed capacity market systems, that gross margin instead comes from annual payments (by the ISO and ICAP market, respectively). From this, one might conclude that there is much more risk associated with the spot energy and operating reserves systems. However, as this year's experience with PJM indicates, ICAP prices can vary significantly, too. In fact, one would expect that capacity prices would vary from near zero to very high as conditions cause the ICAP market to vary from surplus to shortage. Long-term contracts and futures markets are likely to diminish those risks for ICAP market participants, but then the same tools can be used to manage risks in energy-only and operating reserves markets (Graves and Read, 1998).

As a final set of analyses, we have also done runs in which different values of VUE^o have been assumed. Different levels could result, for instance, from various different price caps in the energy market. If the price paid for capacity (in the price-based case) or for operating reserves (in that quantity-based model) is appropriately adjusted, precisely the same LOLP and capacity mix can be achieved as shown in the above table. The effect of a higher VUE^o is to lower the market clearing P_{ICAP}, the required P_{PB}, and either the required operating reserve margin RM_{OR} and/or the price for such reserves P_{OR}. For instance, if the price cap on energy sales is raised to $5000/MWh from $1000/MWh, then the required capacity prices in all cases would fall by about 15%. The total revenue received by generating units in all cases is the same; all that differs is whether it comes out of the energy market or from other sources. The average costs to the

[47] As a comparison, the gross margin from the capacity market in PJM in 1999 was approximately $20,000/MW, while the gross margin from energy sales was approximately $60,000 for a peaking unit with a $45/MWh cost (PJM Market Monitoring Unit, 2000g). Thus, the actual capacity payment is significantly less than the ICAP payment in Table 4.5, while the actual margin from energy sales was considerably greater. However, the total gross margin is approximately correct. The actual distribution of gross margin seems to more closely resemble a combination of the operating reserves and ICAP markets, in which significant revenue comes from both energy when capacity is insufficient to meet reserve requirements and from ICAP payments themselves.

consumer are the same.[48,49] Similarly, if the energy price cap was tightened to, say, \$500/MWh, then the price paid for operating reserves or installed capacity would have to increase in compensation to maintain the same LOLP.

4.4.3. Volatility Results

From the probabilities shown in Tables 4.3 and 4.4 price regime, volatilities can be calculated. One simple index of volatility might be the fraction of the time that the price spikes to VUE^o. By assumption, this probability is the same for the three systems in Table 4.4 that achieve the target LOLP. That is, each has the same expected hours per year in which capacity falls short of demand. However, if we instead consider the size of the price spikes, we see that the systems in Table 4.4 have far lower volatility than the pure spot price system in which the value of unserved energy approaches the true VUE* of \$27,563. In the latter case (Table 4.3), prices are over an order of magnitude higher during the spikes. Thus, the price-based and quantity-based system can both dampen price spikes without damaging reliability, if designed correctly.

However, this conclusion is misleading, because it disregards the volatility of ICAP prices. This volatility can be high, as recent history shows in PJM. In theory, due to random year-to-year variations in loads or capacity, available capacity credits will sometimes be higher than what the market requires, and sometimes lower. In the former case, ICAP prices will hit the ICAP price cap, and in the latter case, ICAP prices should crash to zero (although the actual distribution of ICAP prices is not that extreme). In equilibrium, risk-neutral generators will add capacity until the expected ICAP price (the probability that P_{ICAP} equals the price cap, times P_{ICAP}) equals the price calculated in our model. The probability that P_{ICAP} equals the cap can therefore be calculated, knowing the model's price and the price cap. However, the volatility over the year cannot be estimated by such a calculation, and would require a dynamic stochastic analysis (in the manner of Ilic *et al.*, 2001).

4.4.4. Retirement Decisions for Existing Plants

We now examine the decision to retire or to continue to operate an existing power plant under the alternative capacity market designs. In this analysis, we assume that the plant is relatively small compared to the size of the market, so that its decisions do not significantly affect the prices. The question addressed is: do any of the capacity market designs have a bias in favor of or against retirement of existing plants. This question is important because it has been argued that ICAP systems are biased towards keeping existing plants in service that would otherwise be retired.

---

[48] This is contrary to assertions that are sometimes made that ICAP or other capacity market systems unnecessarily increase costs to customers (*e.g.*, Hanger *et al.*, 2000). In our stylized system, ICAP only increases the cost of power if it also increases the reliability of the system.

[49] It is possible for the energy price to rise so high (here, to above \$27,563) that the ICAP capacity constraint would no longer be binding in the complementarity condition given in §4.2. As a result, P_{ICAP} will fall to zero as more capacity will be provided than required by the ICAP constraint. Peter Cramton (quoted in Hanger *et al.*, 2000) states that the competitive price of ICAP should be zero in general. What we show here, however, is that this can sometimes occur if sufficient incentive to add capacity comes out of the energy or operating reserves market. But if the ICAP constraint is binding (which can happen if the energy and reserve markets are subject to the market failures described in §2), then the ICAP price is generally nonzero.

We answer the question by comparing the economics of keeping a plant in service for the four market designs, assuming as we have above that each is designed to yield the same reliability level. Thus, we are considering the spot market solution from Row 1 of Table 4.3 with the ICAP, operating reserve, and price-based capacity markets of Table 4.4. The starting point is an existing generating plant with marginal cost MC_e \$/MWh, forced outage rate FOR_e, and a fixed cost per unit of capacity of F_e \$/MW/yr that can be avoided if the plant is retired. Assuming risk neutrality, the plant is retired if its expected revenues from the energy, capacity, and operating reserves markets are less than the operating and avoidable fixed cost.

Table 4.7 shows different generating units representing various combinations of MC_e and F_e that result in zero profit under the spot market equilibrium in the top row of Table 4.3. For simplicity, we assume that the existing plant cannot provide operating reserves. Generators with higher marginal costs might be viewed as representing older, less efficient facilities. A forced outage rate of 0.12 is assumed in every case. Thus, these units are at the breakeven point under the spot pricing system. If the argument that ICAP systems increase the attractiveness of keeping existing generators running is true, then we would expect that at least some of the same units would earn positive profits under the ICAP system. However, Table 4.7 contradicts this; when profits are zero under the spot pricing system, they are also zero under any of the other three market designs that yield the same reliability.[50] Thus, the ICAP system does not necessarily make retirement less attractive.

Table 4.7: Profits of Existing Generating Units Under Alternative Capacity Market Designs

Unit	MC_e (\$/MWh)	Fe (\$/MW/yr)	Profit under Energy Spot Market Only (Table 4.3, Row 1) (\$/MW/yr)	Profit under ICAP (Table 4.4) (\$/MW/yr)	Profit under Price-Based (Table 4.4) (\$/MW/yr)	Profit under Operating Reserves (Table 4.4, Last Row) (\$/MW/yr)
1	30	119,590	0	-12	-17	-88
2	35	99,179	0	-12	-17	-87
3	40	78,768	0	-12	-17	-87
4	45	58,358	0	-12	-17	-87

If on the other hand we were to compare the economics of an existing unit under one of the other spot market solutions in Table 4.3 with the ICAP system, it would indeed be true that a breakeven unit under spot pricing would earn positive profits under ICAP. However, this is also true with new generating units, and so there is no discrimination in favor of keeping existing plants. The reason why the plants are more profitable under the ICAP system is that the ICAP payments are greater than the amount of revenue earned during periods of price spikes (Regime 1) in the spot pricing system.

[50] The differences shown in Table 4.7 are actually nonzero, but only because of numerical error in the solutions.

4.4.5. *Prices Spike Before Demand Reaches Available Capacity*

In real markets, bids for energy in the daily auctions are not constant in price for every MW of capacity supplied by each firm. The bids start from a minimum that covers firm's costs and ramp up for the last MW of their available capacity in order to reserve some capacity for the balancing market in case prices are high there. In light of this behavior, prices for energy reflect the scarcity of capacity before the actual deficiency happens. This assumption is included in the simulation by modifying Regime 1 described in §4.1 where shortage-based prices are defined as occurring when the system is at capacity shortage. Here, this threshold is moved 1000 MW below the level at which an actual capacity shortage occurs. The results are shown in the following Tables 4.8 and 4.9, which update the results from Tables 4.3 and 4.4, respectively.

The results for the Pure Spot Market in Table 4.8 show that to obtain a LOLP of 1 in 10 years, the VUE has to be of $12,000/MWh, 43% of the corresponding value observed in Table 4.3 where prices spike occur only when demand exceeds available capacity. By obtaining this reduction, the prices in the simulations are of the order of magnitude of those observed in real markets. The resulting generation mix is the same as in the case described in Table 4.3.

Table 4.8: Pure Spot Market Results, Assuming Price Spikes Occur when Available Capacity is Less or Equal than Demand + 1000 MW

VUE* [$/MWh]	Peaking Capacity [MW]	Baseload Capacity [MW]	Unforced Reserve [MW]	Unforced Reserve Margin [%]	LOLP [1/10yr]	P(Regime 1): Capacity Shortage	P(Regime 3): P = MC_2	P(Regime 4): P = MC_1
27,563	25,918	30,635	3,057	6.30%	0.40	0.03%	52.93%	47.04%
12,000	24,870	30,635	2,062	4.25%	1.00	0.06%	52.89%	47.04%
10,000	24,624	30,635	1,828	3.77%	1.22	0.08%	52.88%	47.05%
1000	20,562	30,636	(2,031)	-4.19%	17.72	0.79%	52.16%	47.05%
500	18,516	30,636	(3,974)	-8.19%	42.83	1.66%	51.29%	47.05%

Accordingly, the peak prices observed for the different quantity- and price-based systems show a significant reduction relative to Tables 4.3 and 4.4. Also, the ICAP prices in the quantity-based system, and in the price-based system are reduced 5% to $61,000/MW/yr. However, essentially the same generation capacity mixes result as in §4.4.2.

The Operating Reserves cases are also modified by including the described price spiking behavior. As a result, the price for Operating Reserves P_{OR} and the VUE are reduced by 8%, and the same generation mix is achieved. For the Variably Priced Operating Reserves case, the reduction is 46%, but there is a slight difference in generation mix, probably due to numerical approximations.

Table 4.9: Comparison of Alternative Market Designs that Achieve LOLP = 1 [day/10 yrs], Assuming Price Spikes Occur when Available Capacity is Less or Equal to Demand + 1000 MW

Capacity Market Design	Peaker Capacity [MW]	Baseload Capacity [MW]	Unforced Reserve Margin [%]	LOLP [1/10yr]	P(Regime 1): Loss of Load	P(Regime 2): Capacity Shortage	P(Regime 3): $P = MC_2$	P(Regime 4): $P = MC_1$
Pure Energy Price Spike (VUE = \$12,000/MWh)	24,870	30,635	4.25%	1.00	0.06%		52.89%	47.04%
ICAP (P_{ICAP} = \$61,005MW/yr; VUE^o = \$1000/MWh)	24,867	30,636	4.24%	1.00	0.06%		52.89%	47.05%
Price-Based (P_{PB} = \$61,000/MW/yr)	24,865	30,636	4.24%	1.00	0.06%	n/a	52.89%	47.05%
7.5% Operating Reserves (VUE^o = \$1860/MWh; P_{OR} = \$1815/MWh)	24,865	30,635	4.24%	1.00	0.06%	0.35%	52.54%	47.05%
7.5% Operating Reserves Variably Priced (VUE^o = \$2750/MWh; $P_{OR\ max}$ = \$2705/MWh)	24,831	30,671	4.24%	1.00	0.06%	0.35%	52.31%	47.27%

5. CONCLUSIONS

ICAP is a mixed command-and-control and market-based tool in which the ISO sets a capacity goal for the pool, and allows the market to determine how that target will be achieved. The report has reviewed economic, regulatory and technical issues concerning the design and operation of ICAP markets. Alternative market mechanisms for assuring generation adequacy have been reviewed, including a pure energy market, price-based systems (fixed capacity payments), and quantity-based systems (tradable ICAP credits, mandatory acquisition of call-options, and operating reserves markets). Their advantages have been discussed, and a modeling analysis performed to explore whether those systems can be used to achieve the objectives of system adequacy and cost minimization under the simplified assumption of pure competition.

In this section we summarize our answers to several important questions concerning the ICAP market based on those analyses. These questions concern whether ICAP can be an efficient means of ensuring adequacy, or whether it instead it leads to market distortions that justify its modification or elimination in PJM. The particular questions we address are:

1. Is ICAP a fictional product with no inherent value?
2. Does ICAP distort energy prices?
3. Is ICAP insufficiently related to system reliability compared to operating reserves?
4. Is ICAP subject to migration, and so cannot ensure adequacy?
5. Does ICAP distort entry/exit decisions?
6. Does ICAP magnify market power problems?

Our conclusions concerning these issues are justified in more detail in §3 and §4 of this report. In summary, we believe that ICAP markets can be effective tools to manage system adequacy. However, as implemented in PJM, the ICAP market has several problems, most of which can be corrected by altering market rules and penalties. The most important correctable problem is that of capacity "delisting" or migration to neighboring markets when energy prices are higher elsewhere (Stoft, 2000a). A problem that is more difficult to correct within a market that has both ICAP and an energy price cap is that energy prices will not reflect system conditions during periods of tight capacity. This could discourage cost-efficient measures to reduce loads or increase generating unit availability at those times. Correcting that problem would require more fundamental reforms. The alternative of substituting operating reserves markets for ICAP is one widely discussed reform. A less well-known but nevertheless promising proposal is to require Load Serving Entities to obtain call options to cover their forecast peaks.

Further research is recommended to obtain more definitive conclusions on these questions. Additional modeling analyses can investigate the effects of market rules on capacity delisting; efficiency losses resulting from a weakening of price signals concerning scarcity of capacity; the exercise of market power in energy and ICAP markets; the effect of ICAP markets on retirement in a system dominated by older facilities; the interactions between capacity decisions and environmental goals; the effect of construction lead times and price expectations on investment under alternative systems (Ilic *et al.*, 2001), and the desirability of locational ICAP requirements. Surveys of project developers and marketers could assess the extent to which ICAP requirements represent a barrier to entry. Finally, empirical studies comparing price behavior in ICAP and non-ICAP systems could test the conclusions of our model concerning the effects of ICAP upon energy prices.

5.1. Is ICAP a Fictional Product With No Inherent Value?

If the presence of ICAP improves system adequacy, it has value from the perspective of the entire market. ICAP yields this value if otherwise energy prices by themselves would fail to motivate sufficient construction. Energy prices alone can be insufficient either because customer willingness to pay to avoid outages is not reflected in energy prices (due to price caps or absence of real-time price signals to consumers) or because producers are risk averse.

Nonetheless, ICAP credits are indeed an artificial commodity whose demand is created by ISO rules. Owning an ICAP credit does not confer the right to any cash flows independent of the ICAP market itself; its value to the owner derives solely from demonstrating compliance with ISO regulations and avoiding penalties for noncompliance. Because credits have no inherent value, credit markets would not spontaneously spring up in the absence of ISO rules.

Although the value of ICAP credits to purchasers is related to noncompliance penalties imposed by the ISO, the cost of ICAP credits to suppliers is more ambiguous and depends on market rules. In the case of New England, it has been correctly pointed out by Cramton and Lien (2000) that the true short-run cost of ICAP is no more than the modest incremental expense of maintaining the plant so that it can produce power. This is because all power plants in New England could be called to operate when ISO-NE has insufficient generation, even if they do not submit bids to the ICAP market or commit the capacity in the bilateral market. As a result, there is no opportunity cost stemming from foregone sales of capacity outside ISO-NE, because such commitments are not allowed. Thus, when capacity is in excess, as it is now in that region, ICAP prices should be at or close to zero, assuming competitive behavior.

In PJM, however, supplying ICAP does have an opportunity cost. During times when energy prices are higher elsewhere and PJM recalls its ICAP, PJM generators have the option to delist from ICAP, in which case they are then free to sell the energy in another market. Thus, PJM generators will require that ICAP revenues compensate them for foregone sales opportunities in more profitable markets. This opportunity cost will be reflected in ICAP bids.

In contrast to ICAP, an example of a product with inherent value is a call option on energy: it entitles the owner to certain cash flows or energy when energy prices exceed the strike price. The mandatory call option proposal that we discussed above (§3.2.3) would substitute a product with a value independent of ISO rules (call options) for a product with no such value (ICAP). The advantages of this proposal are that the value of the call options can be determined by financial analysis methods, while providing protection against price spikes and ICAP-like incentives for capacity construction. In addition, the mandatory call option proposal would prevent the delisting problem discussed later in these conclusions. Therefore, we recommend that the merits of the mandatory call option proposal be considered by PJM.

5.2. Does ICAP Distort Energy Prices?

The purpose of the ICAP market is to ensure that adequate capacity is available to a market. *If* the ISO can identify the socially efficient level of capacity reserves, then an ICAP system can be designed to achieve at least that level (subject to the caveat below about capacity migration). The question here is: how would an ICAP requirement affect energy prices?

Compared to an unfettered energy market, a combination of an ICAP system and energy price cap would moderate price spikes. Our analysis shows that, in equilibrium, generators receive less revenue from the energy spot market, but make it up in the ICAP market. Thus, the reduction of energy spot prices does not eliminate the signal for adding capacity, which is instead given by the ICAP market prices.

Our model shows that under six important assumptions, the following conclusion can be made. If the ICAP system is designed to yield the same level of system reliability as the unfettered spot market, then total revenues to each generator, the equilibrium capacity mix, and the total cost of supply are the same under the ICAP and spot market systems (although the costs to consumers is distributed differently over the year). The assumptions are:

1. Pure competition (no market power);
2. Suppliers who receive ICAP credits cannot "delist" capacity from the market during peak periods and divert it to other markets;
3. Generators add capacity if expected revenues exceed expected costs (risk neutrality);
4. Capacity additions are instantaneous (no lags);
5. Consumers do not respond to real-time prices; and
6. The probability of a generator being available during peak periods equals the annual forced outage rate used to calculate ICAP credits.

These are, of course, strong assumptions. The questions of market power and delisting are discussed further below. If generators are risk averse then mixes and costs of generation might change because ICAP together with price caps will lower price volatility. If the assumption of instantaneous entry of new plants is relaxed, the reality of lead times and uncertainties means that capacity additions may under- or overshoot optimal levels; Ilic *et al.* show that it is possible that this problem may be more severe under ICAP systems than a pure energy price system.

Assumptions 5 and 6 deserve more discussion. If a significant fraction of load does face and respond to real-time prices, elimination of price spikes will discourage economic efficient demand reductions at those times. This is because the true cost of consumption during peak periods will be averaged over the year in the form of ICAP prices. However, a partial fix of this price distortion might be for LSEs to include ICAP costs in customer demand charges. Also, Hobbs *et al.* (2001b) has pointed out that price-responsive load could be charged prices in excess of the cap upon prices paid to generators during times of shortage, with the margin being rebated to the consumers the following year or used to offset fixed ISO operating costs. Regarding the sixth assumption, if there are measures that generators can take to improve on-peak availability that are unreflected in the forced outage rates used to calculate ICAP, then ICAP's diminishment of peak prices could discourage economically efficient actions of this sort. Efforts to base ICAP credits on generation capacity availability during peak periods could lessen but not eliminate this problem.

Thus, whether the energy price changes caused by ICAP represents economically significant "distortions" depends on the extent to which (1) load is responsive to real-time prices and (2) the calculated ICAP differs from on-peak availability. We recommend that additional modeling be undertaken to assess the possible significance of these distortions. Hirst and Kirby (2000) show, for instance, that even a small amount of price elasticity during peak periods can go a long way towards moderating price spikes.

This conclusion about the effect of ICAP on energy prices should be tested empirically. Prices in several regions with and without ICAP markets should be analyzed. Both mean prices and frequency of spikes should be compared. However, such a comparison will be challenging given the multiple causes of price fluctuations, and the lack of transparent price data in some markets, particularly in those regions lacking an hourly markets.

5.3. Is ICAP Insufficiently Related to System Reliability Compared to Operating Reserves?

It has been argued that ICAP rewards "iron in the ground" and not availability and flexibility of generation during peak periods when capacity is truly needed. However, our modeling analysis shows that under the above five assumptions, an ICAP system with an energy price cap and an operating reserves market can provide the same level of adequacy at the same cost. The different market designs give the same amounts and mixes of peaker and baseload capacity. The differences between the systems are the distribution of revenues (between energy, operating reserves, and ICAP) and how energy price spikes will be managed. An operating reserves market yields more energy and operating reserves revenue than ICAP markets, whose loss is made up in the latter system by ICAP payments.

However, two key assumptions behind this conclusion are no consumer response to real-time prices and the tying of ICAP definitions to actual ability to help meet load during peak periods. Because the prices of energy and operating reserves better reflect actual operating conditions under an operating reserves market, responsive loads and measures focused at improving availability during system peaks will be more attractive under that system. For example, if some customers pay real-time prices, an operating reserves system—with its attendant price spikes due to reserves shortages—will motivate customers to cut back just at those times when the system is at risk. This incentive is removed if only long-term installed capacity is traded, and its cost buried in an "uplift." Also, paying for installed capacity only, irrespective of its ramp rates and other operational characteristics, does not reward efforts to improve plant availability and flexibility during times of system peak. In contrast, operating reserves markets can provide such an incentive.

As we noted in our response to the previous question, adjustments can be made to the ICAP system to respond to this potential problem. However, the response will necessarily be incomplete, because the inherent nature of a market combining ICAP and a price cap is to spread revenues throughout the year that would otherwise be concentrated during price spikes. A more thorough reform (mentioned earlier) involving mandatory call options and lifting of the energy price cap could provide incentives that appropriately reflect operating conditions.

5.4. Is ICAP Subject to Migration, and So Cannot Ensure Adequacy?

It is widely recognized that depending on market conditions in PJM and neighboring markets, it can be advantageous for generators to commit to meet loads outside PJM even during times of capacity shortages within PJM. Capacity can be diverted in this manner by delisting an ICAP resource with a two-day notice. The phenomenon arises from the interaction between regions with different market designs. In particular, if neighboring systems have a higher energy price cap than PJM or no cap at all, there will be a temptation to divert capacity when there are large price differences. The effect of this can be that PJM generators are paid twice for their capacity; once through ICAP payments during most of the year, and a second time from price spikes in neighboring systems. The PJM consumers who pay for ICAP will not receive the capacity precisely when it is most needed.

As discussed in more detail in §3, the PJM Market Monitoring Unit has recommended changes in the ICAP deficiency charge in order to try to prevent the delisting problem, and modifications were filed with FERC (PJM-Interconnection, 2001e). However, any change in the penalty will be *ad hoc* to some extent, as there is no accepted basis for determining the "right" penalty.   More fundamental changes, such as increasing the length of notice or switching to an operating reserves or mandated call option system, may be required, and should be considered.

5.5. Does ICAP Distort Entry/Exit Decisions?

A change in the distribution and level of revenues from energy, capacity, ancillary service, and other markets can in theory alter incentives for adding capacity of different types and for retiring existing plants. In particular, critics have argued that ICAP markets prolong the life of old, inefficient plants that would otherwise be retired because ICAP provides a stream of revenue that those facilities would not receive from just the energy market. These plants also tend to have higher emission rates, so extension of their operating lives could worsen air quality problems.

However, under the modeling assumptions listed above under the second question, our analysis shows that there is no such bias towards existing plants in competitive ICAP markets. Plants that lose money in a pure energy spot market will also lose money in an ICAP market that results in the same level of system reliability. The revenues are the same for any particular plant, just distributed differently between energy and capacity payments.

Another criticism is that existence of the ICAP market adds complexity and thereby deters entry by new producers. For instance, to be certified as a producer in the PJM pool, a company has to submit detailed performance reports to PJM for purposes of calculating ICAP credits, and has to learn the rules for the energy and ICAP markets in order to be able to participate in the auctions and bilateral markets. We have not assessed the significance of these barriers relative to other costs involved in entry; a survey of applicants for new plant licenses could indicate whether the complications posed by the ICAP market have raised barriers.

On the other side, exit/entry problems can also affect new LSEs who wish to join the market. Complaints have been made because of the fact that a large fraction of the cost of power for new LSEs has consisted of capacity payments, and that these payments have been predominantly made to owners of existing plants. Our modeling analysis indicates that under our assumptions, the total cost of power to LSEs (including both energy and capacity payments) should not be affected by the presence of an ICAP market. It would be useful to test this conclusion empirically by examining prices in markets with and without ICAP, assuming market price data were available to do so.

5.6. Does ICAP Magnify Market Power Problems?

An important objection to an ICAP requirement is that capacity markets can be gamed. But the existence of a market for installed capacity will not in general lead to more market power than already exists. Generally, if power markets are concentrated, strategic manipulation of price will be attempted in all markets that larger firms participate in (energy, capacity, transmission, and ancillary services). The presence of ICAP together with an energy price cap may simply move market power around, from the spot energy market to the capacity market. However, if ICAP is primarily traded on a long-term basis, it could diminish aggregate market power by encouraging entry.

However, poorly designed rules for ICAP markets can indeed magnify market power. This is indicated by the experience in the ISOs that have ICAP markets. Three different pools with different ICAP rules have shown different results, depending on the design of the market. New England has seen market power exercised in the monthly ICAP. Very high bids have been submitted for supplying capacity in hopes that those bids will set the price. This occurs even though ICAP is in excess supply and has no opportunity cost (as it cannot be delisted and diverted outside the region). What invites this gaming is that the demand for ICAP is automatically entered into the auction as a perfectly inelastic demand curve. Redesign of the ICAP market to allow submission of price-sensitive demand bids (as is permitted in PJM) would help eliminate this gaming.

On the other hand, high ICAP prices in the NYISO do not appear to result from the exercise of market power. New York has defined geographic submarkets for ICAP based on transmission constraints, such as one for the New York City area. In the summer of 2000, the capacity offered in that area's ICAP market was approximately 25% below the ICAP requirement established by the ISO, and the result of any auction in that circumstance, even under pure competition, is an ICAP price that hits the cap (NYISO, 2000). However, in general, geographically separated submarkets will have higher supplier concentrations and more opportunities for exercising market power. An important question is the tradeoff between a more accurate representation of capacity needs (by defining submarkets based on transmission constraints) and the greater opportunity that segmented markets present for strategic manipulation. Tools are available to address such questions (*e.g.*, Berry *et al.*, 1999).

Lastly, in PJM there are claims of market power based on the illiquidity of the market, concentration of suppliers, and the high ICAP prices that have recently been experienced. The PJM Market Monitoring Unit has concluded, however, that the opportunity cost of ICAP is a plausible explanation of those high prices, as there is a correlation between PJM's ICAP prices and energy prices in neighboring systems. Such a correlation is to be expected because PJM generators (unlike those in ISO-NE) can commit their capacity to other markets where they can at some times get higher energy prices. However, continued monitoring is necessary, because manipulation may still have occurred undetected and is possible in the future.

REFERENCES

Berry, C.A.; Hobbs, B.F.; Meroney, W.A.; O'Neill, R.P.; Stewart, W.R. "Understanding How Market Power Can Arise in Network Competition: A Game Theoretic Approach," Utilities Policy, 8(3), 1999, 139-158

Bhavaraju, M.P. "Maintaining PJM Reliability by an Operating Obligation Method" Public Service Electric & Gas. Presented at the PJM FAWG for Consideration. 2000. www.pjm.com

Billinton, R. "Evaluation of Reliability Worth in an Electric Power System", Reliability Engineering and System Safety, 46(1), 1994, pp. 15-23

Billinton, R.; Allan, R.N. Reliability Evaluation of Engineering Systems, Concepts and Techniques, Plenum Press, NY, 1992

The Brattle Group. "Shortening the NYISO's Installed Capacity Procurement Period: Assessment of Reliability Impacts". Final Version. May 2000. www.nyiso.com

Cameron, L.; Cramton, P. "The Role of the ISO in U.S. Electricity Markets: A Review of Restructuring in California and PJM." The Electricity Journal, 12(4), 1999, 71-81

Caramanis, M. "Investment Decisions and Long-Term Planning under Electricity Spot Pricing." IEEE Transactions on Power Apparatus and Systems, PAS-101(12), 1982, 4640-4648

Chao, H.P.; Peck S.C. "Reliability Management in Competitive Electricity Markets," Journal of Regulatory Economics, 14(2), 1998, 189-200

Chao, H.P.; Wilson, R. "Priority Service—Pricing, Investment, and Market Organization," American Economic Review, 77(5), 1987, 899-916

Cramton, P.; Lien, J. "Eliminating the Flaws in New England's Reserve Markets," Working Paper, University of Maryland, March 2000, www.cramton.umd.edu/auction-papers.htm

Doorman, G., "Capacity Subscription – Securing the Balance between Peak Supply and Demand," Proceedings of Market Design 2001, Methods to Secure Peak Load Capacity on Deregulated Electricity Markets, Saltsjobaden, Sweden, June 7-8, 2001, www.elforsk-marketdesign.net/archives/ papers/doorman.pdf

Earle, R.L.; Hanser, P.Q.; Johnson, W.C.; Reitzes, J.D. "Lessons from the First Year of Competition in the California Electricity Markets," The Electricity Journal, 12, 1999, 57-76

EIA/DOE. Electric Power Annual 1998. Energy Information Administration, 1999. DOE/EIA-0348(98)/1

Felder, F. "Energy Caps and the Proposed Reduction of the Installed Capacity Obligation Procurement Period from Six Months to One Month". KEMA. Memorandum to NYISO. November 5, 2000. www.nyiso.com

FERC. Order No. 888. April 24, 1996

FERC. Order Docket No. Oa97-261-000. Conditionally Accepting Open Access Transmission Tariff And Power Pool Agreements, Conditionally Authorizing Establishment of an Independent System Operator and Disposition of Control over Jurisdictional Facilities, and Denying Rehearing. Issued November 25, 1997

FERC. Order ER00-2005-000. Conditionally Accepting CMS/MSS Congestion Management And Multi-Settlement Systems. ISO New England, Inc. Issued June 28, 2000a

FERC. Docket Nos. EL00-62-005 and EL00-62-013. ISO New England, Inc. Order On Compliance Filing. Issued December 15, 2000b

FERC. “Ensuring Sufficient Capacity Reserves in Today's Energy Markets: Should We? And How Do We?” Study Team Discussion Paper, Washington, DC, September 26, 2001

Goulet, D.; Foster. D.; "Notes from PJM ICAP and EPC Users Group meeting October 15, 1999." Commonwealth of Pennsylvania / Office of Consumer Advocate, 1999

Graves, F.C.; Read, J.A., "Capacity Prices in a Competitive Power Market," Working Paper, The Brattle Group, Cambridge, MA, 1998

Green, R. “Competition in Generation: The Economic Foundations,” Proceedings of the IEEE, 88(2), 2000, 128-139

Gross, G.; Garapic, N.V., McNutt, B., "The Mixture of Normals Approximation Technique for Equivalent Load Duration Curves," IEEE Transactions on Power Systems, PWRS-3(2), 368-374, 1988

Hanger, J.; Rohrbach, J.; Adels, P. "Time for PJM to Dump ICAP Too," Penn Future, 2(13), June 30, 2000

Henney, A. “Contrasts in Restructuring Wholesale Electric Markets: England/Wales, California, and the PJM,” The Electricity Journal, 11(August 1998): 24-42

Hinkel, O. "Key RAA Parameters" June 2000. www.pjm.com/committees/documents/RAA/download/key_raa_parameters_06012000.pdf

Hirst, E.; Hadley, S. “Generation Adequacy: Who Decides?”, The Electricity Journal, 12(Oct. 1999), 11-21

Hirst, E.; Kirby, B.; Hadley, S. “Generation and Transmission Adequacy in a Restructuring U.S. Electricity Industry,” Edison Electric Institute, 1999

Hirst, E.; Kirby, B. “Retail-Load Participation in Competitive Wholesale Electricity Markets”. Consulting in Electric-Industry Restructuring, December 2000

Hobbs, B.F.; Iñón, J.; Kahal, M., “A Comparison of Alternative Approaches to Electricity Capacity Markets with Specific References to the PJM System,” Market Design 2001, Methods to Secure Peak Load Capacity on Deregulated Electricity Markets, Saltsjobaden, Sweden, June 7-8, 2001a, www.elforsk-marketdesign.net/archives/papers/hobbs.pdf

Hobbs, B.F. ; Iñón, J.; Stoft, S. “Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?”, The Electricity Journal, 14, August/Sept. 2001b

Hogan, W.W. “Contract Networks for Electric Power Transmission,” Journal of Regulatory Economics, 4, 1992, 211-242

IEEE-USA. “Ensuring Electric Power Reliability: The Challenges Ahead,” IEEE Reliability Symposium, May 24, 2000, Washington, DC

Ilic, M.; Poonsaeng, V.; Skantze, P. “Dynamic Investment in Electricity Markets and Its Impact on System Reliability” Proceedings of Market Design 2001, Methods to Secure Peak Load Capacity on Deregulated Electricity Markets, Saltsjobaden, Sweden, June 7-8, 2001, www.elforsk-marketdesign.net/archives/papers/mit.pdf

ISO New England Inc. Press Advisory, "ISO New England Mitigates January Installed Capacity Market Price," April 7, 2000a

ISO New England Inc. “Filing to FERC ICAP Termination effective June 1st, 2000”. May 8, 2000b. www.iso-ne.com/market_rules_and_procedures/documents/ MRP_ICAP_Filing/ ICAP-00-May.doc

Jaffe, A.B.; Felder, F.A. “Should Electricity Markets Have a Capacity Requirement? If So, How Should It Be Priced?” The Electricity Journal, 11(10), 1998, 52-60.

Kahn, J.R. The Economic Approach to Environmental and Natural Resources. 2nd ed., The Dryden Press, 1998

Kueck, J.D. *et al.* Load as a Reliability Resource in the Deregulated Market. Oak Ridge National Laboratory, February 2001. Draft

McCormick, J. Strategic Energy L.L.C., Presentation at the “Joint BRX & RAA-RC 7/31/00 Meeting”, Agenda Item #3-C, PJM, July 31, 2000

Michaels, R.J.; Ellig, J. “Electricity: Price Spikes by Design?” Mercatus Center, George Mason University, 1999. www.cienergy.org/calendar_archives/energy99/pdf/Michaels.PDF

Mira J.; Sanchez M.J., "Reducing Computational Costs for Reliability Indicators of Electric Power Generating Systems," Reliability Engineering and System Safety, 67(3), 2000, 233-240

NERC. Reliability Assessment 1998-2007. North American Electric Reliability Council, 1998

NERC. Reliability Assessment 2000-2009. North American Electric Reliability Council, 2000

NYISO. Ancillary Services Manual. New York Independent System Operator, 1999. www.nyiso.com/services/documents/manuals/pdf/ancserv.pdf

NYISO, “Summer 2000 ICAP Strip Auction Results,” 2000 www.nyiso.com/ summer_2000_icap _strip_auction_results.pdf

Oren, S.S. "Capacity Payments and Supply Adequacy in Competitive Electricity Markets," VII Symposium of Specialists in Electric Operational and Expansion Planning, Curitiba, Brazil, May 21-26, 2000

Panzar, J.C.; Sibley, D.S. Public Utility Pricing Under Risk: The Case of Self-Rationing, American Economic Review, 68(5), 1978, 888-895

PJM-Interconnection. "PJM Manual for Accounted-For Obligation Manual M-17", 1997a

PJM-Interconnection. "PJM Manual for Load Data System M-19", 1997b

PJM-Interconnection. "PJM Manual for PJM Reserve Requirements Manual M-20", 1997c

PJM-Interconnection. Market Operation Committee Meeting, Aug. 27, 1997d

PJM-Interconnection. Reliability Assurance Agreement among LSEs in the PJM Control Area. PJM ISO, 1997e, last update, September 2001, *raa.pdf* in www.pjm.com

PJM-Interconnection. FERC filing "Revisions to Operating Agreement, Schedule 11". November 14th, 1998a

PJM-Interconnection. "PJM Manual for Rules and Procedures for Determination Of Generating Capability Manual M-21," 1998b

PJM-Interconnection. "PJM LMP Implementation Training Course," 1998c

PJM-Interconnection, VCMUG - Visible Capacity Market User Group Minutes and Documents, 1998d, www.pjm.com/Committees

PJM-Interconnection. Public Interest & Environmental Organization User Group. Minutes of the 5th Meeting Hershey Hotel, Hershey, PA, May 7, 1999a, www.pjm.com/Committee Documents

PJM-Interconnection. Reliability Assurance Agreement (RAA) Reliability Committee (RC) 15th Meeting, April 27th, 1999b, www.pjm.com/Committee Documents

PJM-Interconnection. "PJM Manual for Billing M-29," June 1st, 2000a

PJM-Interconnection. "PJM Manual for Installed Capacity: Generation Data Systems M-18," 2000b

PJM-Interconnection. "PJM Operating Agreement Accounting Manual M-28," June 2000c

PJM-Interconnection. Market Monitoring Unit (MMU) June 2000 ICAP investigation presented to the Energy Market Committee (EMC), July 5, 2000d, www.pjm.com

PJM-Interconnection. "Short Term Changes to Ensure Adequacy," White Paper in Future Adequacy Working Group-FAWG, November 2000e, www.pjm.com/Committees: FAWG

PJM-Interconnection. FAQ on Installed Capacity and External Transaction Activities (Answers to User Questions), PJM, 2000f

PJM-Interconnection. Market Monitoring Unit, "PJM Interconnection, State of the Market Report 1999," June 2000g

PJM Interconnection. FERC filing "PJM West Reliability Assurance Agreement among Load Serving Entities in the PJM West Region," March 14th, 2001a

PJM Interconnection Members Committee, 26th Meeting Agenda, April 26, 2001b, www.pjm.com/Committee Documents

PJM Interconnection FERC Filing Letter - Amendment to Schedule 11 of the RAA. (regarding Deficiency revenues), March 7, 2001c

PJM Interconnection FERC Filing Letter – Extension of ALM Periods under FAWG Proposal, April 5, 2001d

PJM-Interconnection. FERC filing Requesting Amendment to RAA among LSEs (regarding CDR policies), April 6, 2001e, www.pjm.com /raa_206.pdf

PJM Interconnection. “Minutes of the 38th Meeting,” Reliability Assurance Agreement (RAA) Reliability Committee (RC), April 23, 2001f, www.pjm.com

PJM-Interconnection. “Generator Interconnection Manual -- Communication and data requirements of the PJM and local control centers. (Manual M- 14),” PJM ISO, 2001g

PJM-Interconnection. “PJM – February 2001 Load Forecast Report,” February 2001h, www.pjm.com

PJM-Interconnection. “PJM – Fundamentals of Unforced Capacity & PJM – eCapacity,” PJM Training Documents, Version 2, 2001i, www.pjm.com

PJM-Interconnection. “PJM - Generator Unavailability Subcommittee (GUS) Manual M-22,” PJM ISO, www.pjm.com/training/customer/inscap/downloads/gus_manual.pdf, 2001j

PJM-Interconnection. “PJM eGADS Manual M-3”, Revision 00, 2001k, www.pjm.com

Rau, N. “The Need for Capacity in the Deregulated Electrical Industry - A Review.” In Proceedings of the IEEE Power Engineering Society Winter Meeting, IEEE, 1999

Ruff, L. "Capacity Payment, Security of Supply and Stranded Costs," presentation to the National Spanish Electricity System Commission, Madrid, June 7, 1999 (cited by Oren, 2000)

Schweppe, F.C. "Power Systems 2000". IEEE Spectrum, July 1978

Schweppe, F.C.; Caramanis, M.C.; Bohn, R.E. “Optimal Spot Pricing: Practice and Theory.” IEEE Transactions on Power Apparatus and Systems, PAS-101(9), 1982, 3234-3245

Schweppe, F.C.; Caramanis, M.C.; Tabors, R.D.; Bohn, R.E. Spot Pricing of Electricity. Kluwer Academic Press, Boston, 1987

Sheffrin, A.; Rahimi, F.; Hildebrandt, E.; Casey, K. Report on Impacts of RMR Contracts on Market Performance, California ISO Market Surveillance Unit, 1999

Shepherd, W.G.. “Electric Power: From Monopoly to Effective Competition?”, in L. Deutsch, ed., Industrial Studies, Second Edition, M.E. Sharpe Publishers, New York, 201-215. 1998

Singh, H.; Jacobs, J.. Capacity Products and ISO Markets. PG&E Energy Services, 2000. Draft

Singh, H. "Call Options for Energy: A Market Based Alternative to ICAP" Paper presented at the PJM-FAWG Committee Meeting October 2000. www.pjm.com/committees/ user_group/… /20001102_pge_call_opt_detail.pdf

Soder, L. “Who Should Be Responsible for Generation Capacity Addition?”, presented at the International Conference on Electric Utility Deregulation and Restructuring, and Power Technologies (DRPT 2000), April 4-7, 2000, City University, London, UK. www.ekc.kth.se/ees/lennart_drpt2000.html

Sotkiewicz., P. “Market Design in the NYISO: Issues and Performance.” In NEMS/AEO 2000 Conferences in Washington DC, Energy Information Agency, 2000

Stoft, S. PJM's Capacity Market in a Price Spike World, PJM Market Monitoring Unit, University of California Energy Institute POWER Seminar, March 2000a

Stoft, S., Power Economics; Designing Markets for Electricity, Chapter 7, "Capacity Requirements," Draft, Version March 16th, 2000b. www.stoft.com

Strategic Energy. “Directed ICAP: To Improve Reliability of America’s Power Supplies”, Presented at the ICAP Working Group meeting, September 24, 2001. http://www4.nyiso.com/services/documents/groups/bic_icap_group /meeting_materials.html

Uhr, C.W. “Driving Peak Load Management Technologies with Coincident-Peak Pricing.” Proceedings of Market Design 2001, Methods to Secure Peak Load Capacity on Deregulated Electricity Markets, Saltsjobaden, Sweden, June 7-8, 2001, www.elforsk-marketdesign.net/archives/papers/uhr.pdf

Vázquez, C., Rivier, M., and Perez-Arriaga, I. “A Market Approach to Long-term Security of Supply,” Proceedings of Market Design 2001, Methods to Secure Peak Load Capacity on Deregulated Electricity Markets, Saltsjobaden, Sweden, June 7-8, 2001, www.elforsk-marketdesign.net/archives/ papers/vazquez.pdf

Vickrey, W. "Responsive Pricing of Public Utility Services," Bell J. of Economics and Management Science, 2(1), 337-346, 1971

Virginia State Corporation Commission - Staff-Report. Developments in Wholesale Power Market. Virginia State Corporation Commission, 1997. Case #PUE950089. www.state.va.us/scc/news/wholesale.htm

Wolak, F.A. “Market Design and Price Behavior in Restructured Electricity Markets: An International Comparison,” POWER Working Paper Series - University of California Energy Institute PWP-051. 1997

Wolak, F.A.; Green, R.; Bushnell, J.; University of California Energy Institute. “Design Choices in the Organization of Electricity Markets,” 1998, www.stanford.edu/~wolak

Wu, D.J.; Kleindorfer, P.; Zhang, J.E. "Optimal Bidding and Contracting Strategies in the Deregulated Electric Power Market." Working Paper, Department of Operations and Information Management, The Wharton School, University of Pennsylvania

APPENDICES

Appendix I. Additional Definitions

Accounted-for Obligation: This is an Obligation based on load ownership and PJM pool reserve requirements. This can result in purchases and sales of unforced Capacity. The Accounted for Obligation for each Party is equal to the LSE Obligation, across all zones, over a Planning Period, determined on a daily basis, summed monthly for billing purposes. The principal tool used in establishing the final LSE Obligation is the web based eCapacity Application.

Active Load Management (ALM): Active Load Management applies to interruptible customers whose load can be interrupted at the request of the PJM OI. Such a PJM OI request is considered an Emergency action and is implemented prior to a voltage reduction.

Capacity Benefit Margin: Portion, in MW, of transmission import capability that is set aside for capacity import purposes in PJM OI's calculation of the amount of required generating capacity necessary to provide reliable service. This is determined by PJM through PJM import capability studies.

Capacity Externality: Externality associated with maintaining sufficient reserve capacity to prevent loss of service, or outages.

Control Area: An electric power system or combination of electric power systems to which a common automatic generation control scheme is applied in order to:

> (1) match, at all times, the power output of the generators within the electric power system(s) and capacity and energy purchased from entities outside the electric power system(s), with the load within the electric power system(s);
>
> (2) maintain scheduled interchange with other Control Areas, within the limits of Good Utility Practice;
>
> (3) maintain the frequency of the electric power system(s) within reasonable limits in accordance with Good Utility Practice; and
>
> (4) provide sufficient generating capacity to maintain operating reserves in accordance with Good Utility Practice.

Diversified Peaks: The Diversified Peaks for the PJM zones are calculated based on the PJM weather normalized actual peak, diversity factor. Adjustments are made for Summer and Winter peaking LSEs.

Diversity Factor(DF): A five-year rolling average value expressed in per-unit, quantifying seasonal (Summer to Winter) peak load shape for a given zone.

Electric Distribution Company or Zonal Entity (EDC): *see* Local Distribution Company

Externalities: Costs or benefits borne by third parties that are not included in the monetary cost of a product (*e.g.*, air pollution caused by electricity generation). More specifically, *technical* externalities are costs or benefits arising from the impact of a decision upon a third party's utility function or production function. *Pecuniary* externalities arise instead from changes in market prices caused by a decision.

Forecast Zone Requirements: Individual zonal requirements based on Forecast Pool Requirements and zonal load values.

GADS: The Generating Availability Data System GADS is, since 1982, an information service that has helped the electric utility industry find solutions to a costly problem: generating unit unavailability. Analysis of the GADS database reveals historical trends from which availability improvement strategies can be developed. With data collected for more than 75,000 unit-years of unit and equipment failure data, GADS is valued because of its data consistency and quality. 182 generating facility operators in the United States and Canada voluntarily participate in GADS. These facility operators represent 92% of the installed capacity in the United States and Canada, representing 95% of the fossil-steam and 99% of the nuclear capacity

LMP: The Locational Marginal Pricing is a method to take into account the energy price differences from bus to bus in an electricity network. The prices in each bus are equal to the marginal cost of supply to that bus (Hogan, 1992)

Load Aggregator (LA): A licensed entity that may provide (sell) energy to retail customers within the service territory of a Local Distribution Company. Also known as Electric Generation Supplier (EGS).

Load Serving Entity (LSE): Any entity (or the duly designated agent of such an entity), including a load aggregator or power marketer, (I) serving end-users within the PJM Control Area, and (ii) that has been granted the authority or has an obligation pursuant to state or local law, regulation or franchise to sell electric energy to end-users located within the PJM Control Area. EDCs and LAs are types of LSEs.

Local Distribution Company (LDC): A company in whose service territory Load Aggregators are providing energy to retail customers and whose distribution system is being used to transport the energy. Also known as Electric Distribution Company (EDC).

LOLP: The probability of capacity shortage P is the expected fraction of time (or hours/year) that system generation capacity is less than total demand. By "expected", we mean the probability-weighted average. This is what PJM refers to as its "loss of load probability", and is also known as a loss of load expectation (LOLE). The expected frequency of outage F is the expected frequency of occurrence of outage events (*e.g.*, per year). Some users of the term LOLP actual refers to a frequency (*e.g.*, "one outage occurring every 10 years, on average"). PJM's target of 1 day in 10 years literally means a probability of 0.000274 = (1 day)/(365 days * 10 years) (R. Gramlich, formerly at PJM, personal communication). For the reliability target in our models in §4, we assume that the 1 day in 10 year LOLP standard of the PJM Reserves Requirement Manual is desired (PJM-Interconnection, 1997d), which we interpret as 24 hours out of 87,600 hours

Marginal Unserved Energy: The amount of unserved energy that is reduced by adding an increment of capacity.

Maximum Generation Emergency: Emergency declared by the PJM OI in which it anticipates requesting one or more Capacity Resources to operate at its maximum net or gross electrical power output, subject to the equipment stress limits for such Capacity Resource, in order to manage, alleviate, or end the Emergency. The following is a table showing the emergency operating procedures of PJM. After the Maximum Generation Emergency has been declared, and if the system is not balanced, curtailments start to be applied.

Emergency Operating Procedures

Emergency Generation Alert - To provide an early alert that system's conditions may require the use of the PJM emergency procedures. It is implemented when Maximum Emergency Generation is called into the operating capacity.

Primary Reserve Alert - When the estimated operating reserve capacity is less than the forecast primary reserve requirement.

Voltage Reduction Alert - When the estimated operating reserve capacity is less than the forecast spinning reserve requirement.

Voluntary Customer Load Curtailment Alert - Whenever the estimated operating reserve capacity indicates a probable future need for voluntary customer load curtailment. No customers are contacted.

Primary Reserve Warning - When available primary reserve capacity is less than the primary reserve requirement but greater than the spinning reserve requirement, after all available secondary reserve capacity (except restricted maximum emergency capacity) is brought to a primary reserve status.

Voltage Reduction Warning & Reduction of Non-Critical Plant Load - When available spinning reserve capacity is less than the spinning reserve requirement, after all available secondary and primary reserve capacity (except restricted maximum emergency capacity) is brought to a spinning reserve status.

Manual Load Dump Warning - When available primary reserve capacity is less than the largest operating generator or the loss of a transmission facility jeopardizes reliable operation after all other possible measures are taken to increase reserve.

Maximum Schedule Generation - When the need for generation is greater than the highest incremental cost curve.

Maximum Emergency Generation - When members have been ordered to Maximum Schedule Generation and a need for more generation still exists.

Load Management Curtailment - The purpose of Load Management Curtailment Steps 1-4 is to provide additional load relief.

Voltage Reduction & Curtailment of Non-Essential Building Load - When load relief is still needed to maintain tie schedules or to relieve transmission constraints.

Voluntary Customer Load Curtailment - When load relief is still needed. Advance notice is required.

Radio & TV Load Curtailment Appeal - When previous steps are not expected to provide sufficient relief.

Manual Load Dump - Implemented when all other relief measures have been ineffective.

Metered Entity: A Local Distribution Company within the PJM Control Area that provides distribution and metering services to customers in its territory.

Planning Period: “Planning Period” shall initially mean the 12 months beginning June 1 and extending through May 31 of the following year, or such other period established by the Reliability Committee established under the Reliability Assurance Agreement.

PJM Office of the Interconnection: The PJM OI consists of the facilities and staff engaged in the implementation of the PJM Operating Agreement and administration of the Tariff.

Power Exchange Market: Institutions dealing with the commercial side of power provision. Selling offers and buying petitions are cleared time-ahead of the actual provision and consumption of power in the system. Some power systems have the Power Exchange separated from the ISO (CA), and others assign the function to the same ISO (PJM)

Reserve Margin: The percent of generating capacity needed above the expected maximum demand. The unforced reserve margin considers only expected available capacity (equal to the sum of 1-FOR times the capacity of each unit, where FOR is the forced outage rate of the unit).

Retail Customer: The energy end- user; interfaces only with the LSE and EDC, not with PJM.

Retail Load Responsibility: The agreed-upon hourly load, within the service territory of the Local Distribution Company, for which the Load Aggregator must provide energy to customers.

Retail Transaction: An energy transaction scheduled between a Load Aggregator and a Local Distribution Company for the Load Aggregator to supply energy for retail load in the LDC's service area.

Summer Peaking Zone: A system whose maximum one hour load during the period of June through September exceeds its reduced winter peak.

Unserved Energy: The amount of energy demanded at a certain price that cannot be supplied with available capacity.

Value of Unserved Energy: The total monetary value per unit energy (MWh) that consumers are willing to pay to avoid an outage, or some other measure of damage due to outages. These costs can be divided into three categories: rationing costs, disruption costs, and lost consumer surplus. Rationing costs are incurred by the utility in determining how to allocate electricity when a scarcity of supply occurs. Disruption costs are the direct costs to the consumers of a shortage. Lost surplus costs express the amount the consumer would be willing to pay if electricity were available (which can include disruption costs).

Winter Peaking Zone: A system whose reduced winter peak is greater than its maximum one hour load during the period of June through September.

WTP: Consumer Willingness to Pay is the maximum amount of money a consumer is willing to pay in order to receive or acquire a specified good. This is a measure of the benefit of that good to the consumer.

Zone: An area within the PJM Control Area, as set forth in the PJM Open Access Tariff and the Reliability Assurance Agreement. Schedule 16 of the RAA defines the ten distinct zones that comprise the PJM control area.

Appendix II. PJM Reserves Planning Process

PJM OI enforces on the system the obligations set in the Reliability Agreement to maintain short-run and long-run generation capacity levels. The short-run levels are obligations to the PJM OI, which buy the following necessary services from generators in a hour basis:

- Spinning Reserve (10 minute synchronized) = Largest Unit or 900 MW
- Primary Reserve or Supplemental (10 minute non-synchronized) = 1700 MW
- Secondary Reserve (30 minutes non- synchronized) = 2500- 4500 MW, based on system conditions.

The long-run reserve capacity level establishes the magnitude of capacity required, above peak load, to ensure the reliability of the PJM control area. This appendix describes how PJM calculates ICAP requirements for the LSEs. PJM OI follows a planning process that determines the capacity that each LSE in the pool needs to supply its forecasted load for each planning period (June 1st to May 31st following year), two years in advance of the actual working period (Reliability Assurance Agreement, PJM-Interconnection, 1997e). The objective is to achieve a Loss of Load Probability (LOLP) of less than 1 day in 10 years in adherence to the North-American Reliability Council criteria. The requirements are established in a procedure involving several sub-committees and technical groups (Load Analysis (LAS), Generator Unavailability (GUS), Load and Capacity (L&CS), Planning Committee) and later approved by the RAA-Reliability Committee.

The process of determining Capacity Obligations is set in the Reliability Assurance Agreement (RAA). But the terms, procedures, and responsibilities are explained in detailed in a series of Manuals that PJM Interconnection produced. Among them are the following:

- "Reliability Assurance Agreement Among Load Serving Entities in the PJM Control Area" (PJM-Interconnection, 1997e, last update, September 2001);
- “Generator Interconnection Manual -- Communication and data requirements of the PJM and local control centers. (Manual M- 14)”
- "PJM Manual for Capacity Obligations Manual M-17" (former *Accounted-For Obligation*). (PJM-Interconnection, 1997a). Describes the methods for calculating Capacity Obligations and Peak Season Maintenance obligations;
- "PJM Manual for Load Data System M-19" (PJM-Interconnection, 1997b). Describes the input requirements process, computer programs and reports produced with the stored load data;
- "PJM Manual for PJM Reserve Requirements Manual M-20", (PJM-Interconnection, 1997c);
- "PJM Manual for Rules and Procedures for Determination Of Generating Capability Manual M-21", (PJM-Interconnection, 1998b). A set of rules and procedures for the determination of generating capability;

- “PJM - Generator Unavailability Subcommittee (GUS) Manual M- 22.” A set of procedures to project generator unavailability for PJM allocation studies and data required for specific planning applications (PJM 2001j);
- “PJM eGADS Manual M-23” (PJM-Interconnection,2001k); and
- "PJM Billing Manual M-29" (June 1st, PJM-Interconnection, 2000a).

The manual "Installed Capacity: Generation Data Systems M-18" was retired on December 1, 2000 (PJM-Interconnection, 2000b) and all references to it in other manuals should be ignored. The manuals refer to the Pool-Wide Choice Date, which is set as 01/01/2000. On this date several calculations changed their expressions, and in what follows we consider the post-Pool-Wide Choice Date references.

The determination of the Capacity, or Accounted-For, obligations for each LSE is made in two steps:

- First, a yearly process in which PJM OI evaluates the control area performance and makes projections into the future about the load and capacity requirements.
- The second process is a daily process in which each LSE is notified of its load responsibility.

Control Area Projections

The goal of this part of the process is to set several important system parameters: Installed Reserve Margin (IRM), an Average Resource Performance (EFORd) and a Forecast Pool Requirement (FPR). The process uses probabilistic methods to take in account the uncertainties inherent to data based on forecasts. The data that the process requires and the parameters it calculates are summarized in the following table.

Inputs and Outputs of the Accounted-For Obligation Calculation.

INPUTS		*OUTPUTS*
Load forecast Resource performance Resource plans ALM Regional interface capabilities Target LOLP	=> *Probabilistic Methods* =>	IRM Avg. Resource Performance EFORd Forecast Pool Requirement FPR

The input information includes both load forecasts and supply forecasts for each LSE. It also takes in account the LOLP that the pool sets as its target operation target. The terms are explained as follow.

Load forecast

The **Load forecast** in PJM and other control areas is the expected load demand for future years at system peak, expressed in MW. The forecast evaluates internal load, and tries to determine which external conditions might put the system's reliability at risk.

For these forecasts, PJM OI gathers information from many sources and analysis it with several committees. The PJM control area is divided into 10 sub-regions, each having a Electric Distribution Co., or LDC. The LDC companies, are also known as Wires companies, own the distribution wires in each area (see Figure II.1 for a map of the sub-regions). The companies feed PJM OI with hourly load data in its wired region to form a database, called the PJM Daily Load & Capacity file (DCL). These data are used in PJM OI's load studies, and they are also distributed monthly as information items to the LDCs.

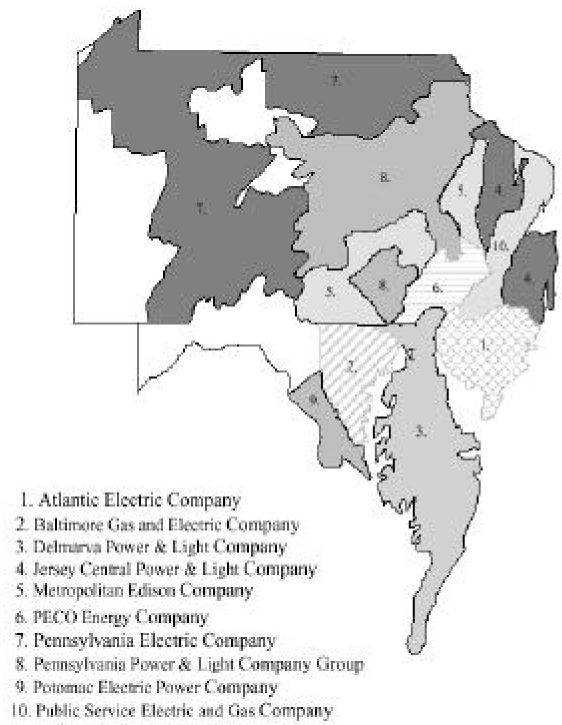


Figure II.1: PJM Transmission Zones

The LDCs also submit to the PJM OI the forecasted loads, peak loads and available monthly ALM in their zone of influence for the following 10 years, calculating and aggregating the information supplied by the LSEs. It contains the following information:

- seasonal peak load forecasts for each Planning Period as provided by each LSE reflecting (a) load forecasts with a 50 percent probability of being too high or too low, (b) summer peak diversities determined by the PJM OI from recent experience, and (c) ALM;
- forecasts of aggregate seasonal load shape of the LSEs that must be consistent with forecast averages of 52 weekly peak loads prepared by the LSEs in each LDC's systems; and
- variability of loads within each week, due to weather and other recurring and random factors, as determined by the PJM OI.

With this information, the Load Analysis Subcommittee (LAS) also performs several studies. First, it puts together the pool's load forecast for the following 10 years, which it formalizes in the February Load Report. The 2001 Report can be obtained in the PJM Website (PJM Interconnection, 2001h). Figure II.2 shows the PJM peak load projections for the next 10 years as published in the report. It assumes an average 1.5%/year growth rate for the Summer peak for the period, and an average 1.1% growth rate for the Winter peak.

Second, it develops a weather normalization which establishes a relation between weather conditions and the PJM daily peak for a season. For this purpose, it considers two periods, Summer (June, July, and August) and Winter (December, January, and February). For each period it determines the PJM maximum unrestricted hourly load (Peak) (adding estimates of Active Load Management), and records weather conditions in three airports within the area (Lehigh Valley International (ABE), Philadelphia International (PHL), and Washington National (DCA)). For the summer period, the afternoon peak is registered, and for the winter season, the evening peak is noted. These weather conditions are transformed into demand-related indexes (Temperature-Humidity for the summer period, and wind-speed adjusted temperature, for the winter period). A linear least-squares regression is used to describe the relation peak-load versus weather conditions, which result in a slope and an intercept. Next, the weather indexes for the maximum season peak for the last 20 years are averaged (20 year rolling average) and with the slope and the intercept, the weather-normalized peak (W/N Peak) is calculated for each season. The PJM OI brings the W/N Peaks to the RAA-Reliability Committee by its October meeting for approval or review.

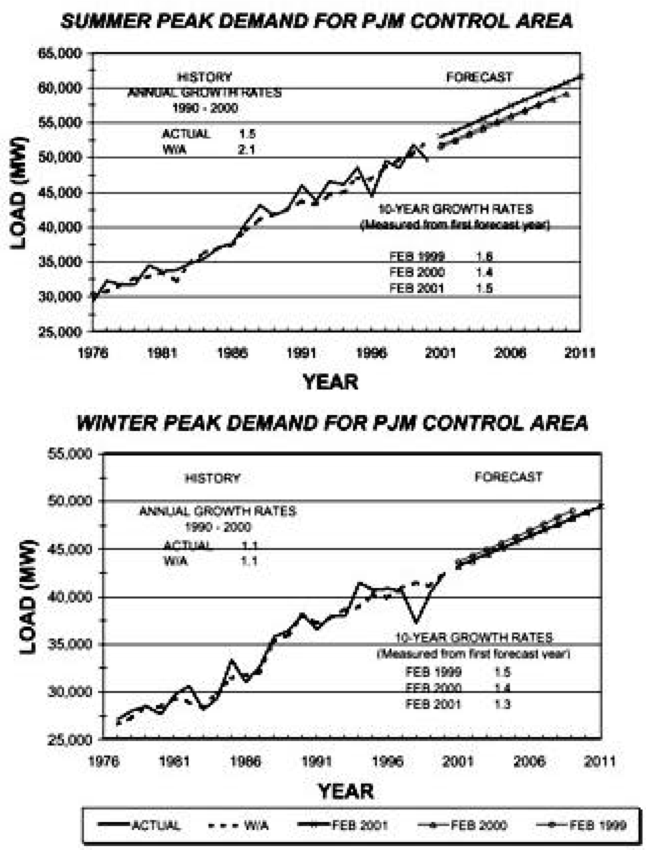


Figure II.2: Summer and Winter PJM Control Area Peak Demand Forecast

Third, using the actual load registered in each LDC's zone, the winter and summer W/N Peaks are allocated to each zone in proportion to its shares of the PJM actual peaks. This is done with a method called the five Coincident Peaks (5CP). Adding the load and ALM data for each hour in each season, the unrestricted loads are created. The five PJM unrestricted highest daily peaks (CP) are identified and are used in the allocation procedure. Each peak is segregated zone determining how each zone contributed to the peak. The zone's contributions to each CP are averaged for each of the 8 zones (GPU is treated as one zone). With this averaged peak load contribution, each zone's share can be determined.

Finally, an allocation of the peaks among the different zones is made. The season's W/N peak for that season is multiplied by each share to determine how the weather-normalized peak is to be distributed to the LDC's zones. Table II.1 shows the winter and summer seasons shares of W/N peaks among the geographical zones in PJM, and compares the forecasted values for peak in each season and the W/N values for the corresponding season. As it stands now, there is a lag of one year in the information being used, *e.g.*, the values of W/N peak for the winter 01/02 peak share is the 99/00 winter W/N. This procedure was proposed to be changed but the modifications were not approved as of this writing (PJM Interconnection, 2000e).

Table II.1: Winter and Summer Peak Shares for the 8 Zones in PJM (GPU treated as one zone)

	99/00 Winter Peak Share (%)	Winter W/N Peak and Zone Shares [MW]	2000 Summer Peak Share (%)	Summer W/N Peak and Zone Shares [MW]	Relation Between Forecasted & W/N Peak
PJM	100.00	42,580	100.00	52,350	
PSE&G	15.53	6,613	18.48	9,673	
PECO	14.26	6,072	14.65	7,670	
PLGroup	15.99	6,809	12.33	6,454	
BGE	12.77	5,437	12.03	6,296	
GPU	19.80	8,431	19.59	10,255	
PEPCO	10.86	4,624	11.43	5,984	
AE	3.90	1,661	4.58	2,400	
DPL	6.89	2,934	6.91	3,619	
2001 PJM summer forecasted peak [MW]				52,930	1.011
2001/02 PJM winter forecasted peak [MW]		43,762			1.028

Resource Performance Information

Resource performance information [dimensionless] consists of a set of historical outage rates. This information is self-reported by generation owners to the NERC/GADS system. GADS is a database used for entering, storing, and reporting generating unit data concerning outages, unit performance, and fuel performance. Data is entered and stored on an individual-unit, individual-event basis. Data that is entered by each LSE is accessible only to that LSE and to authorized PJM OI personnel. GADS event (outage) data are used by OI to feed mathematical models to calculate outage rates, to prepare data for NERC, and to perform special studies. These data include minimum information (name, location, type, fuel, nameplate rating), performance data (monthly generation, service hours-SH, fuel consumption, etc.), and outage event data (record of times and causes for unit being out of service, Forced Outage Hours-FOH). The outage events are classified as Planned, Maintenance, and Forced, and may be full or partial. Being unplanned, forced outages represent a degree of reliability of a given generator, and have the greatest impact on reliability in the planning process. These forced outages could be declared if energy is exported and the unit cannot be recalled by PJM. Two values are calculated based on outage data for each generator:

Forced Outage Rate FOR (%) = FOH/ (FOH + SH) * 100

Equivalent Forced Outage Rate EFOR (%) = (FOH+EFDH) / (FOH+SH+EFDH+RS) * 100

The EFDH is defined as Equivalent Forced Derated Hours, SH is the service hours, and RS is the Reserve Shutdown. FOH includes outages hours in which a generator is declared in a forced outage condition when it does not deliver the energy required by the ISO under Maximum Generation Emergency.

Resource Plans

Each LSE has to evaluate how is the forecasted load going to be met, either by existing or planned generation assets.

Active Load Management

ALM gives credit derived from specific customer information and provided by LSEs. It represents the anticipated reduction of customer Peak Load Contribution. The ALM Factor is a weighting factor for Active Load Management that represents probabilistic value of ALM contribution at time of peak.

Regional Interface Capabilities

As PJM accepts generation units outside its control area to be considered as capacity resources for its studies, it has to evaluate consider transfer capabilities to provide for interregional assistance by those resources.

The modeling gives the following products:

- **Installed Reserve Margin (IRM)** is the required total system reserve capacity (over and above peak demand) within PJM. It is the percent margin of the forecast PJM Control Area requirement for the planning period over the coincident forecast planning period peak of all LSEs. Other terms for this concept are PJM RM, PJM Control Area RM**,** and Pool RM.
- Average resource performance statistics expressed as an **Equivalent Forced Outage Rate Demand (EFORd)** for all PJM capacity resources. The average EFORd shall be the average of the capacity-weighted EFORds of all units committed to serve load in the PJM Control Area. It characterizes individual units and is calculated with data from the NERC/GADS system, and is calculated for the generator as

  EFORd = EFOR x ***f***

  being ***f*** a parameter that adjusts EFOR to reflect on/off peak demand and predicts the probability of failure to operate when needed. EFORd provides the best measure of ability to supply load at the PJM peak. It is based on a 12 months rolling average. The EFORd of a unit in service five or more full calendar years at the time of forecast shall be the average rate experienced by such unit during the five most recent calendar years.
- A **Forecast Pool Requirement (FPR)** for capacity calculated from the IRM and EFORd values. It is a weighting factor to convert load based information to an equivalent requirement or unforced capacity based values:

  FPR (%) = (1 + IRM) * (1 - EFORd) *100.

The values approved by the RAA committee for the Planning Periods is shown in the following Table II.2 (PJM-Interconnection, 2001f). During the April 2001 RAA-Reliability Committee meeting the values for the planning period 2003/2004 were approved.

Table II.2: System Capacity Requirement Parameters.

Planning Period	IRM	Avg. EFORd	FPR	ALM FACTOR
1999/2000	20.00%	9.52%	1.0858	0.967
2000/2001	19.50%	9.76%	1.0784	0.987
2001/2002	19.00%	9.52%	1.0767	0.965
2002/2003	19.00%	8.43%	1.0897	0.966
2003/2004	18.00%	7.18%	1.0953	0.954

Accounted-for Obligation

The Accounted-for Obligation value is the amount of capacity resources that an LSE should acquire. The zonal weather-normalized peaks that result from the PJM Load Forecast are used by to determine the Capacity Obligations for each LSE operating in the corresponding LDC's distribution zones (PJM-Interconnection, 1997b). The PJM OI has the load data for each LSE, and therefore daily sets the value calculating on the variations that the peak contribution may present (new customers, or loss of consumers). The obligations is set 36 hours in advance of their load responsibility to allow the Parties to manage the resources in advance (PJM Inerconnection, Training 2001i). The Accounted-For Obligation daily requirement is calculated by the following formula:

$$\text{Accounted-For Obligation for each LSE} = [(\text{Peak Load Contribution})_{LSE} - (\text{ALM Credit})_{LSE} * (\text{ALMFactor})] * (\text{FPR}) / 100$$

Peak Load Contribution is the daily summation of the weather-adjusted peak loads for the end-users for which the LSE was responsible on that billing day, as determined by the procedures set forth in the PJM Manuals. The determination of the Peak Load Contribution depends on the month of the season in which the requirement is made. The two possibilities are as follows:

1. For each billing month during a Planning Period up to and including the month in which the peak load for the PJM Control Area during the summer period is known, *the Peak Load Contribution is the daily summation of the* ***weather-adjusted actual coincident summer peak for the previous summer*** *of the end-users for which the LSE was responsible on that billing day.*
2. Beginning in the billing month following the month in which the peak load for the PJM Control Area during the summer period is known, *the Peak Load Contribution of the LSE shall be determined for the remaining months of the Planning Period on a daily basis*

__considering the daily summation of the weather-adjusted actual summer peak load__ of the end-users for which the LSE was responsible on that billing day.

The **ALM credit** is the amount of Active Load Management credit that the LSE is certified to own.

In order to comply with the obligations, the LSEs have to show owned resources, acquire credits under direct bilateral transactions, purchase ICAP credits in the PJM Capacity Credit Market, or take credits for certified ALM programs. The Capacity Credit Market operates on a daily and monthly basis. Each LSE's own resources are derated according to the resources' past availability rates. These derated values are credited for meeting the capacity requirements. For a generator with a given nameplate installed capacity, the systems considers that that plant is able to deliver energy at the Unforced Capacity rate of

Unforced Capacity = [Installed Capacity] * (1 - EFORd)

This capacity should be "deliverable" and with an approved physical capability of the transmission network as the procedures for establishing deliverability of capacity resources sets in the RAA's Schedule 10. Otherwise, in the event of emergency, these resources might not be available to the network due to transmission constraints.

The PJM Interconnection Office has a monthly billing process that tracks the LSE's installed capacity on a daily basis during the actual planning period. The Accounted-For Obligation is then used to determine if an LSE is deficient in capacity, in which case it will incur deficiency charges pursuant to the Operating Agreement of PJM Interconnection "Distribution of Deficiency Charges" (Schedule 11, RAA).

Appendix III. Capacity Deficiency and Recall

The mechanism to verify the commitment to meet LSE obligations is managed within the PJM eCapacity system. This Capacity Credit Market clears daily and monthly commitment transactions. In comparison, NY-ISO operates on a six-month seasonal commitment. A study has been under way to analyze the impact of the reduction of this period to 1 month or increasing it to 1 year (Brattle Group, 2000).

Each LSE's own resources are derated according to the resources past availability rates. For a generator with a given nameplate installed capacity, the systems considers that that plant is able to deliver energy at the Unforced Capacity rate of:

Unforced Capacity = [Installed Capacity] * (1 - EFORd)

Capacity Credits are the entitlement to a specified number of megawatts of Unforced Capacity for the purpose of satisfying capacity obligations imposed under this Agreement and that are acquired by a Party through bilateral purchase or pursuant to Schedule 11 of the Operating Agreement, or any successor schedule. (Reliability Assurance Agreement; PJM-Interconnection, 1997e). Those derated capacity values are credited to each LSE.

The LSE can collect revenues from its installed capacity through the offer of the energy in the energy market, in bilateral energy transactions, in the operating reserves ISO purchases, or in the real-time balancing market (operating reserves credits). The capacity should also be "deliverable" in the sense that it should have enough Available Transfer Capability (ATC). Otherwise, in the event of emergency, these resources might not be available to the network due to transmission constraints.

Credits for Operating Reserves are calculated for each of the following situations (PJM Operating Agreement Accounting Manual, 2000c):

1. pool-scheduled generating resources (day-ahead and balancing markets);
2. pool-scheduled synchronous condensers (balancing market);
3. pool-scheduled transactions (day-ahead and balancing markets);
4. canceled pool-scheduled resources (balancing market);
5. resources providing quick start reserves (balancing market); and
6. resources reduced for reliability purposes (balancing market).

If applicable, when a resource is started during the day at the direction of the PJM OI, the resource's real-time offer amount for that day includes its startup costs based on the appropriate hot, intermediate, or cold state of the resource.

The operating reserve costs are distributed to the generators proportionally to the day-ahead demand plus cleared day-ahead exports, and proportionally to the real-time deviation from day-ahead schedules. Operating Reserves were purchased during May and June 2000 at an average rate of 0.66181 [$/MWh], and a maximum rate of 1.0617 [$/MWh] (www.pjm.com/pub/account/operrsvsrates/). The PJM OI calculates the resource's hourly balancing energy market value as the difference between the scheduled energy that the plant would deliver and the actual real-time energy delivered in that day, times the real-time Local Marginal Price for that transmission zone.

If the generator cannot cover its Account-For Obligations with its Unforced Capacity and Capacity Credits acquired in the ICAP market, the ISO will charge the Capacity Deficiency Rate (CDR) for the amount of the daily deficiency in MW. This CDR charge is actually a price cap for Installed Capacity deficiency.

CDR is set annually by the Reliability Committee as a surrogate for an Installed Capacity resource (currently $160/MW-day). It is expressed and calculated as an Unforced Capacity equivalent:

$$\$160 / (1\text{-Pool wide average EFORd}) = \$160 / (1 - 0.0952) = \$176.83$$

It was initially set at $58.4/kW-yr (= 58.4/kW * (1000kW/Mw) / (365day/yr) = $160/MW-day), "the cost of installing a combustion turbine generator" (Henney, 1998). The CDR on a given day is doubled if the Unforced Capacity committed to the pool is less than the total Pool Accounted-For Obligation. The CDR escalates for an individual participant if it fails to meet at least 95% of their obligation for more than 30 days in a rolling one-year period (RAA, PJM-Interconnection, 1997f). The value for CDR is thought to be insufficient penalty for keeping PJM capacity in the Pool, as expressed at the Reliability Assurance Agreement (RAA) Reliability Committee (RC) 15th Meeting minutes (PJM-Interconnection, 1999b) and by Stoft (2000a). The ISO filed with FERC to modify the level and conditions for the CDR (PJM-Interconnection, 2001e).

During the operations day, a given generator that commits its capacity to the Capacity Market assures that its generating power will be available in case of emergency. A Declaration of Emergency is issued when there are insufficient resources to meet load after the market-clearing price has hit the tariff ceiling (Market Operations Committee; PJM-Interconnection, 1997d).

If PJM declares Maximum Generation Emergency (PJM - FAQ on ICAP; PJM, 2000f), it will recall resources. Stoft (2000a) describes this procedure as follows.

- The ISO first takes all day-ahead energy bids to cover for the missing supply.
- If the bids are not sufficient, the ISO can call the operating reserves.
- If the available reserves are not sufficient, ISO calls on all the resources committed in the ICAP market.

If an owner commits the capacity to the Pool, sells the power to external market, and he is recalled to the pool, the ISO curtails the external sales on the emergency. If there is abuse of the rules related to capacity commitment to PJM, the Market Monitoring Unit will conduct an investigation with potential penalties, and in the case the generator tries to avoid its commitment, the ISO will charge 365 times the CDR for that day in which it was unable to deliver the energy (RAA-Schedule 14, PJM-Interconnect, 1997f).

Appendix IV. Miscellaneous PJM Operating Parameters and Capacity Price Data

For different planning periods. the market operating parameters that PJM is using are summarized in the following table (Hinkel, 2000).

Table IV.1. Assumed Reliability Factors, PJM

Planning Period	IRM	Avg. EFORd	FPR	ALM FACTOR
1999/2000	20.0%	9.52%	108.58	0.967
2000/2001	19.5%	9.76%	107.84	0.987
2001/2002	19.0%	9.52%	107.67	0.965
2002/2003	19.0%	8.43%	108.97	0.966

The following table summarizes PJM's working capacity and load characteristic values for this summer 2000.

Table IV.2. PJM Capacity Data (Sources: (1) PJM-Interconnection, 1997e; (2) Hinkel, 2000)

Parameters [units]		Source
Capacity Deficiency Rate (Installed) [$/MWDay]	$160.00	2
Capacity Deficiency Rate (Unforced) [$/MWDay]	$177.30	2
Pool Short CDR (Unforced, $/MWDay) (2 * $177.3) []	$354.60	2
Installed Reserve Margin (IRM) []	19.50%	2
Average EFORd (April 2000 PJM average) []	6.42%	1
5-Year Average EFOR d []	9.76%	2
Forecast Pool Requirement (1 + IRM) * (1 - 5Year EFORd) []	1.0784	2
Net Capacity Resources (2000 MAAC EIA 411) [MW]	58,277	1
Net Unforced Capacity [MW] (Net Capacity * (1-EFORd))	54,535	1
ALM (2000 MAAC EIA 411) []	1,881	1
ALM Factor []	0.987	2
ALM Nominated as of June 6, 2000 [MW]	1,686	2
Unforced Value of ALM (ALM * ALM Factor * FPR) [MW]	1,795	2
Coincident Peak CP [MW]	50,510	2
Unforced Obligation (CP * FPR, without ALM) [MW]	54,470	1
ALM Offset to Obligation (Net) [MW]	2,002	1
Net Unforced Obligation [MW] (Unforced Obligation - ALM Offset)	52,468	1
Net Unforced Capacity Obligation	52,676	2
Capacity Excess (Deficiency) [MW] (Net Unforced Capacity - Net Unforced Obligation)	2,068	1

The PJM ICAP market has been increasing the amount of capacity traded since 1998, and the prices experienced have been relatively low most of the time. The following chart shows ICAP price trends during 1999-2001 (Figure IV.1). Prices surpassed the CDR charge several days in 2000 and 2001, and then settled at an average of 177$/MWh. The price excursion in June 2000 was investigated by the Market Monitoring Unit (PJM Interconnection, 2000d). The occurrence of high prices since January 2001 motivated the ISO to submit three filings to FERC regarding modification of ICAP rules (PJM Interconnection, 2001c, 2001d and 2001e). One filing concerns the extension of the ALM credits beyond the May 31 deadline, and the other concerns the CDR charge and the distribution of revenues obtained by the ISO as a result of that charge.

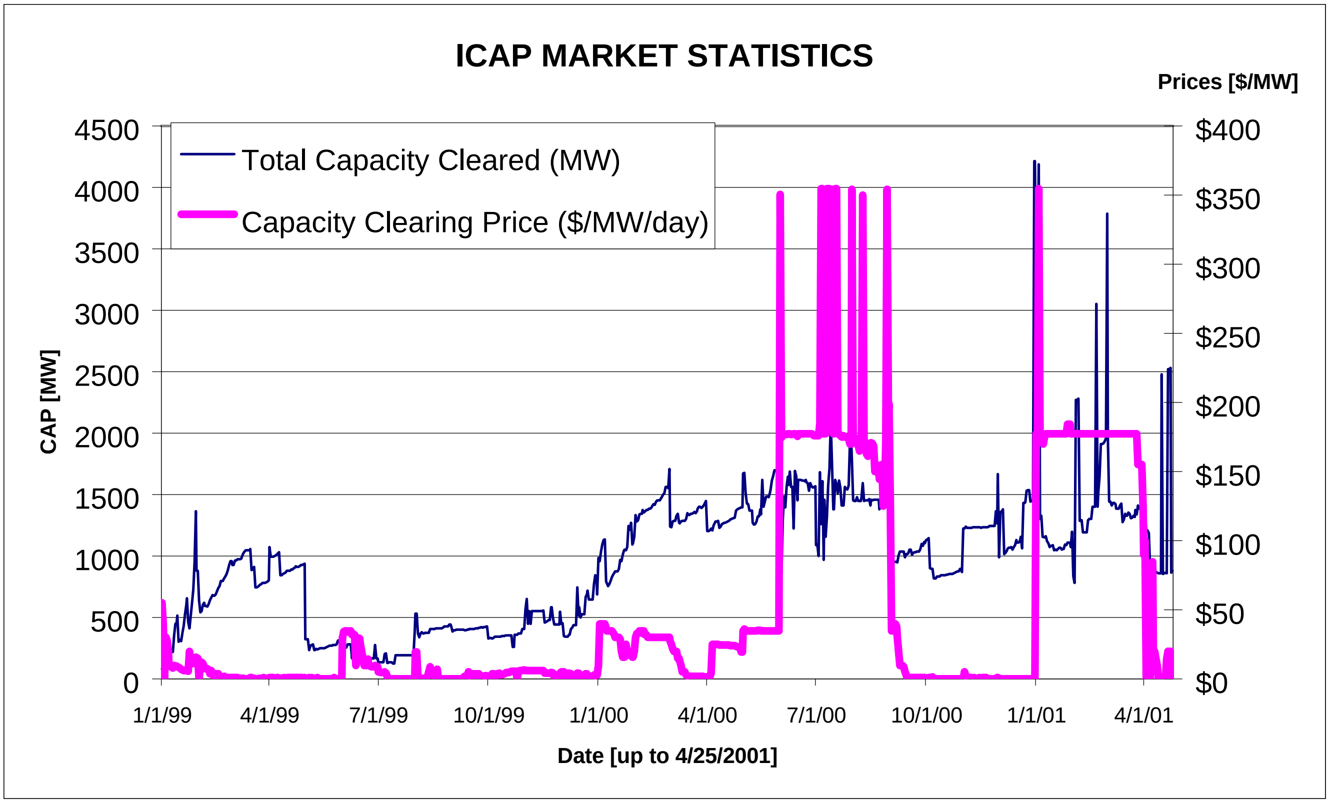
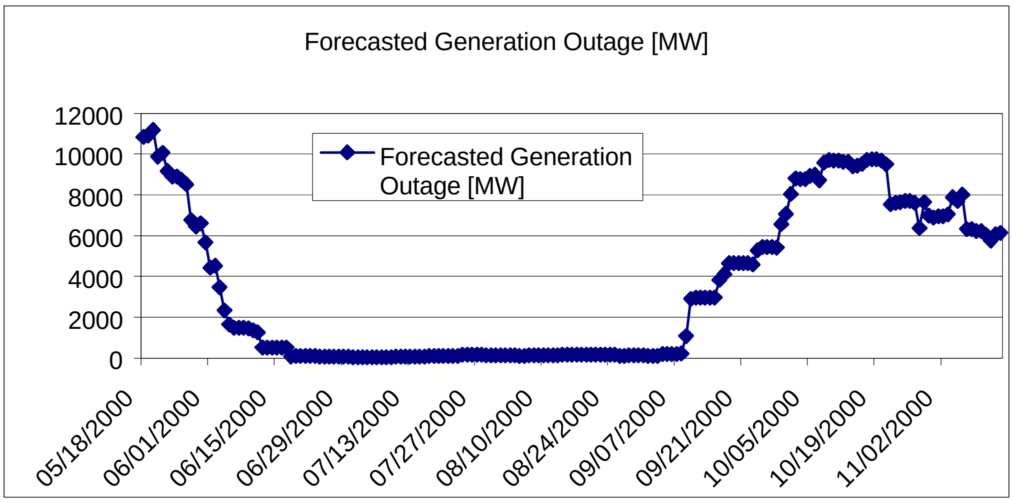


Figure IV.1. Installed Capacity Market Clearing price. [$/MW/day] (www.pjm.com/Historical Daily Capacity Credit Market Results)

The forecasted generation on planned outage for PJM is shown in Figure IV.2. For instance, it was forecasted that for June 18, 2000 and June 11, 2001 that the generation outage in MW would be minimal. For the June 2000 event, Figure IV.2 shows that the pool had almost full available capacity, so the ICAP market clearing price should have had a low value reflecting the non-binding capacity constraint, as in previous months. In light of this expectation, the fact that the price hit the ceiling of the CDR charge, and peaked to the Pool Short CDR, was surprising. Several issues regarding ICAP market may be acting at the same time, *i.e.*, imperfect competition (market power—generator withholding of capacity), imperfect information (the risk for forced outage being greater than calculated), and transaction costs (low market liquidity).

However, this expectation of low ICAP prices during times of adequate capacity assumes that prices for energy outside PJM were not so high as to tempt ICAP providers to delist their capacity. Singh and Jacobs (2000) argue that the fluctuation in ICAP prices results from capacity being diverted regions outside PJM. They suggest that since the installed capacity requirement does not significantly vary from day to day (it depends on the annual peak not the daily peak), its price should not differ that much over the year. However, in the summer, the capacity becomes valuable elsewhere, and so its price can be driven up.

These issues are discussed further in Appendix V, which focuses on the June 2000 ICAP price excursion.



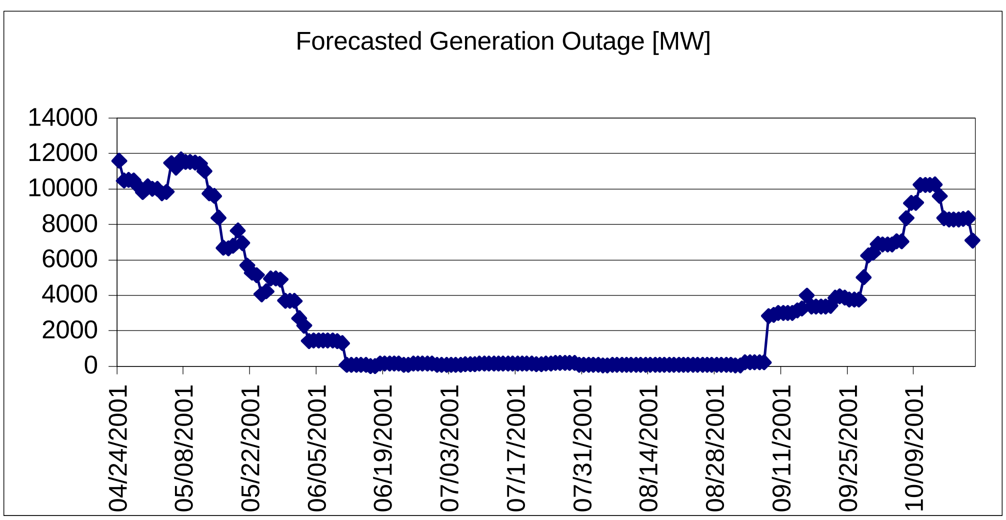


Figure IV.2. Forecasted Generation Outage [MW] for Two Periods in 2000-2001 (ftp://www.pjm.com/pub/account/genoutages/)

Appendix V. June 2000 PJM Capacity Market MMU Report Summary

On June 1, 2000, prices in the PJM daily capacity credit markets reached the highest level since this market was introduced in late 1998. Daily capacity market prices fell on June 2, but remained high by historical standards for the balance of June. In response to these prices, various members of PJM requested that the Market Monitoring Unit investigate to determine whether market power or market manipulation was the cause of the price increases.

The Market Monitoring Unit (MMU) performed an investigation and, based on the results of that investigation, provided a presentation to the Energy Market Committee (EMC) on July 5, 2000 (PJM-Interconnection, 2000d).

Prices in the PJM daily capacity market rose on June 1, 2000, and remained high through the end of June as a result of underlying economic fundamentals. These economic fundamentals include, at the most basic level, the level and price sensitivity of demand for capacity and the level and price sensitivity of supply of capacity. There is no reason to believe that market power explains the high prices in the June daily capacity markets or that the daily capacity market prices in June were increased by the unilateral action of a market participant or the joint action of a group of market participants.

The conclusion reached by the MMU and presented to the EMC was that prices and behavior in the capacity markets for June appeared to be consistent with the underlying supply and demand fundamentals and that there was no evidence of market manipulation.

Figure V.1 shows the prices and quantities traded in daily and monthly capacity markets from January 1999 through June 2000. In 1999, capacity market prices averaged $52.86/MW-day over all capacity markets including daily, monthly, and multi-monthly markets. Monthly capacity market prices averaged $70.66/MW-day. Daily capacity market prices averaged $3.63/MW-day, while the highest daily market price was $55/MW-Day.

Figures V.2 and V.3 show energy forward prices for PJM and NYISO, in June and July-August respectively. The two figures show an increased spread in the energy price between the two regions, which resulted in a significant increase in delisted capacity in PJM. Figure V.4 shows the evolution of the capacity market demand bids, supply offers, and capacity market prices for May and June 2000. Comparing the market conditions in May with market conditions in June, it is observed that the gap between the MW level of demand offers and supply offers narrowed in June, having a situation for one day in which the demand was higher than the offer. At that time, ICAP prices spiked to twice the amount of the CDR..

Figure V.5 shows capacity exports and the resulting pool capacity position for the year 2000 prior to and including the early June events. Delisted capacity increased from 876 MW on May 31, 2000 to 2,031 MW of installed capacity on June 1, 2000. To put this in context, the average delisted capacity was 906 MW in 1999 and the maximum delisted capacity was 1,776 MW in 1999. The delisted capacity on June 1, 2000 exceeded the maximum delisted capacity in 1999 by 255 MW while the delisted capacity on other June days was less than the maximum level of delisted capacity in 1999. The decrease in supply in the daily capacity credit markets led to an increase in price. In response to this price increase, 552 MW of capacity returned to PJM as capacity resources over the five days after June 1st.

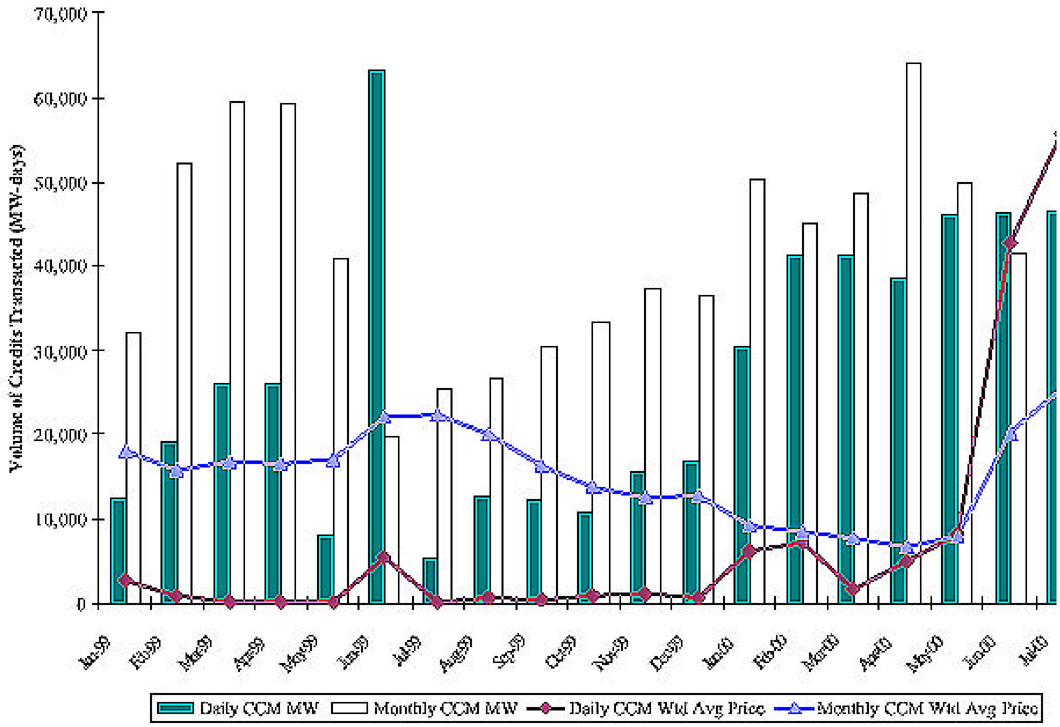


Figure V.1. 1999 & 2000 Daily vs. Monthly Capacity Credit Market (PJM-Interconnection, 2000d)

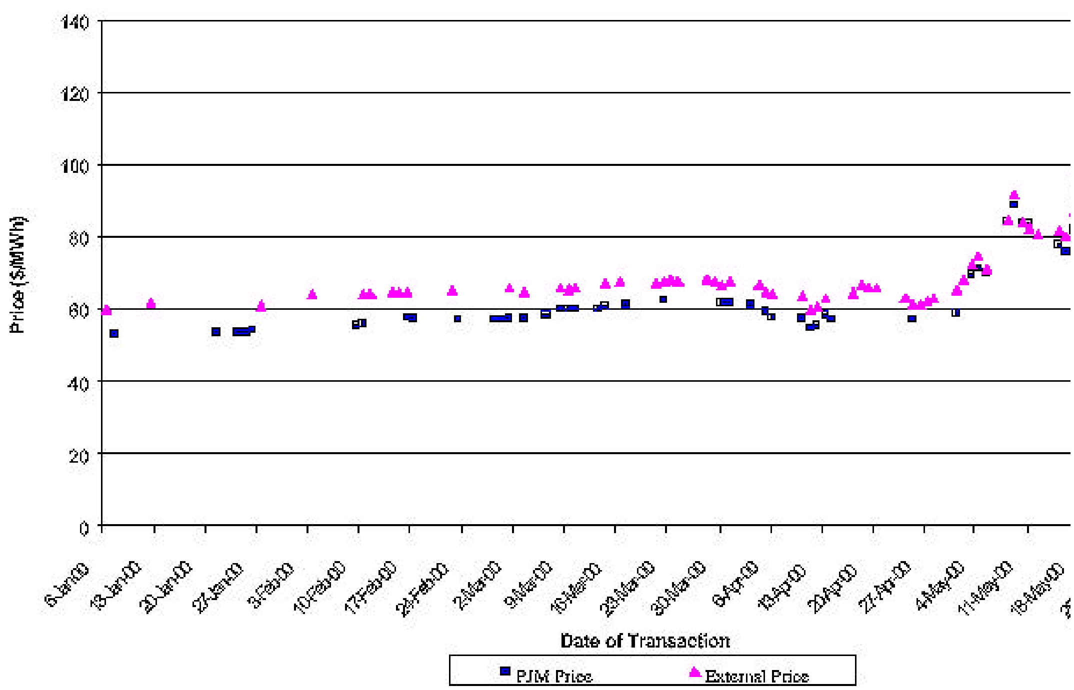


Figure V.2. June 2000 Forward energy contract price, PJM and NYISO (PJM-Interconnection, 2000d)

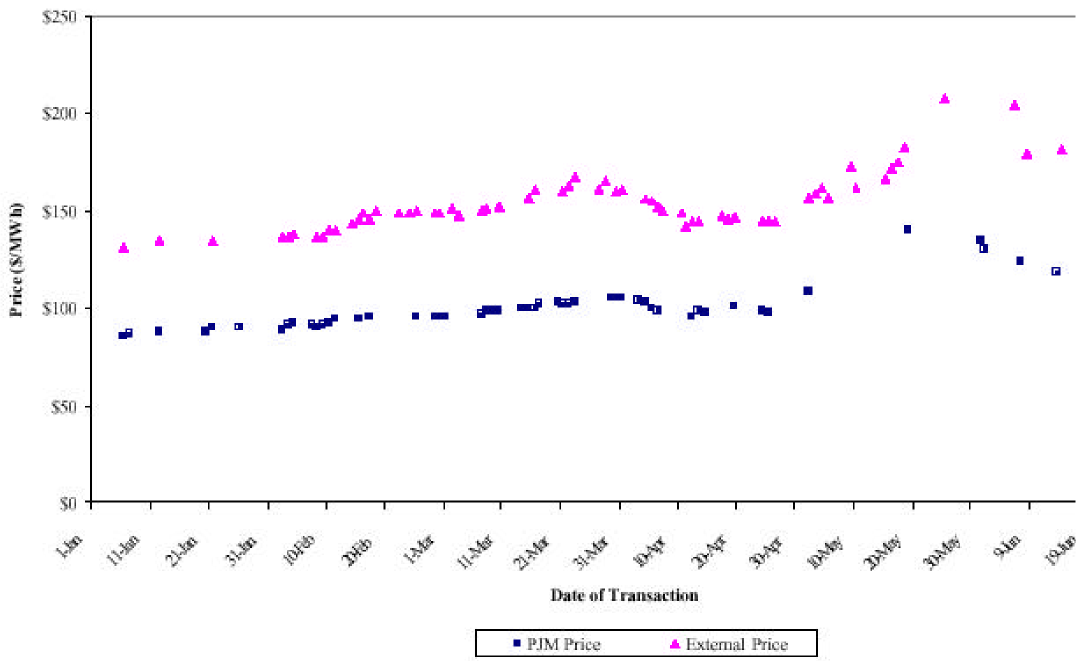


Figure V.3. July-August 2000 Forward Energy Contract Price (PJM-Interconnection, 2000d)

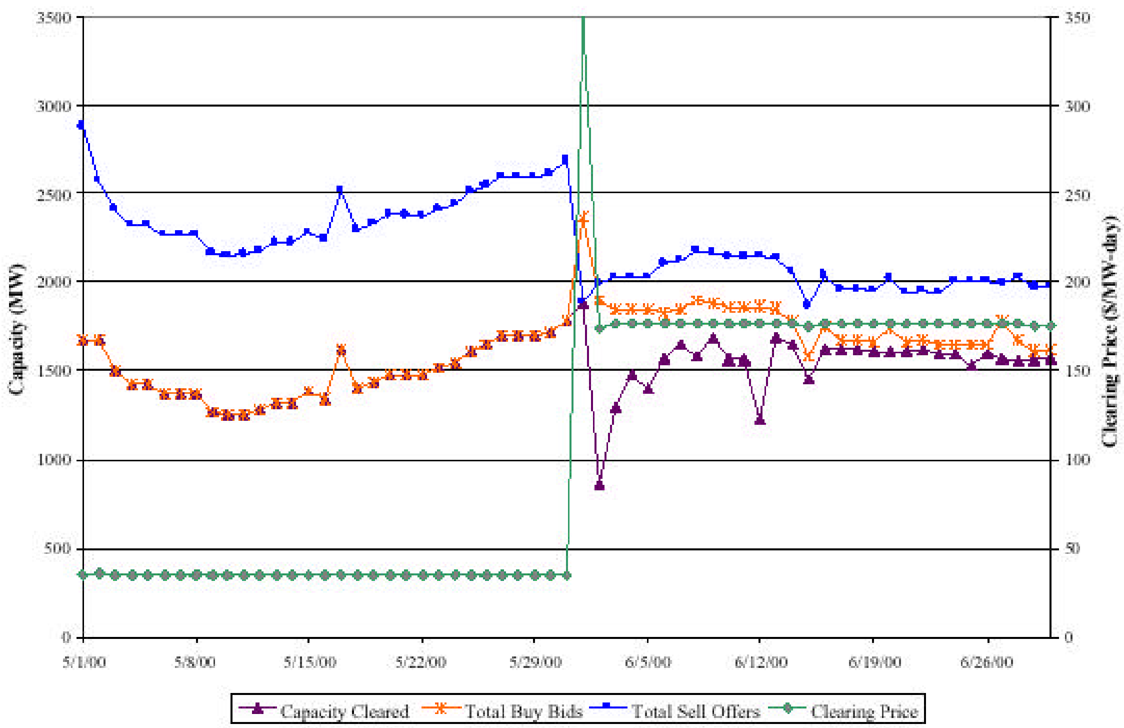


Figure V.4. Daily Capacity Credit Market Summary

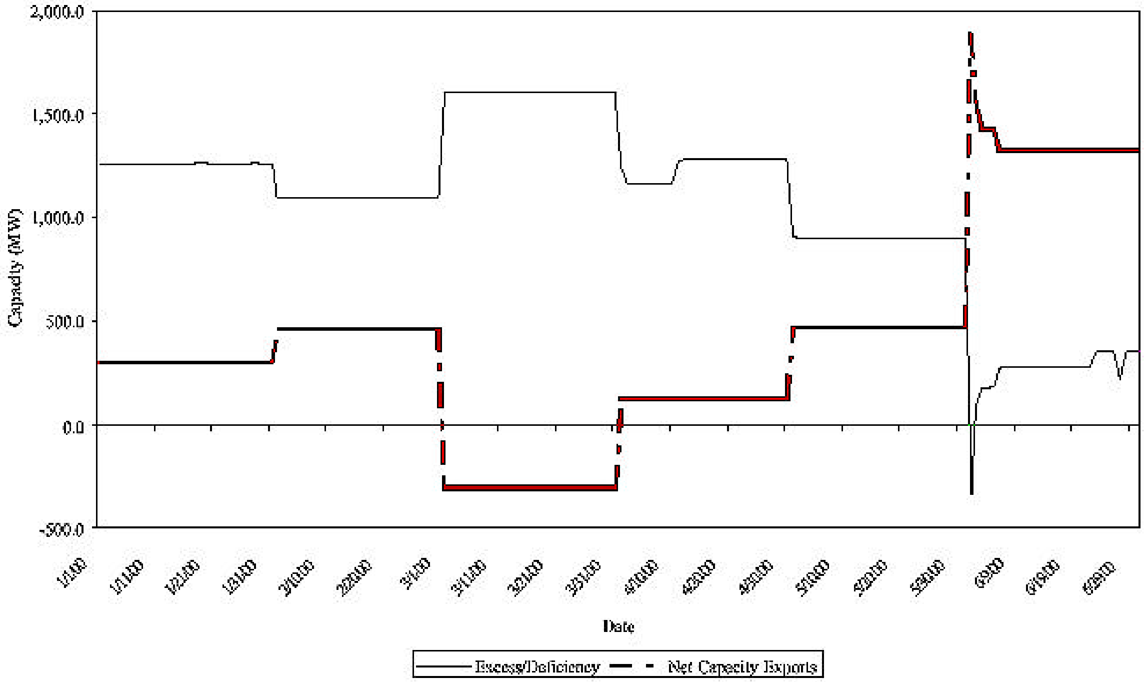


Figure V.5. External Capacity Transactions and Pool Excess/Deficiency (PJM-Interconnection, 2000d)