Els UK Job: CDI ch09-i047172 16-11-2007 4:15p.m. Page:325 Trim:165×240MM Float:Top/Bot TS: Integra, India

Part III
Capacity, Resource Adequacy, and Investment

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T. W:30PC open recto 1 Color 49 Lines

Els UK Job: CDI ch09-i047172 16-11-2007 4:15p.m. Page:327 Trim:165×240MM Float:Top/Bot TS: Integra, India

Chapter 9
Resource Adequacy
Alternate Perspectives and Divergent Paths

PARVIZ ADIB,[1] ERIC SCHUBERT,[1] AND SHMUEL OREN[2]

[1]Automated Power Exchange, Santa Clara, California, USA and BP North America Gas and Power, Houston, Texas, USA; [2]University of California, Berkeley, USA

Summary

Ensuring that deregulated wholesale markets meet resource adequacy targets has been a concern of policymakers for more than a decade. This chapter reviews the difficult and controversial evolution of capacity resource adequacy mechanisms in the United States. Next, the alternative used in Texas is examined: an energy-only resource adequacy mechanism based on the successful Australian model. The necessary conditions for a sustainable energy-only approach and a potential transition mechanism to an energy-only approach are also discussed.

9.1. Introduction

While electric industry restructuring has been increasingly refined across the world for about 15 years, there is still a significant debate on whether direct capacity remuneration should be an integral part of competitive wholesale electricity markets. This chapter will address this question. The authors present a selected historical account of resource adequacy mechanisms throughout the world with a focus on developments in the United States, where the capacity framework has been going through substantial evolution and policy debate in multiple markets under the Federal Energy Regulatory Commission (FERC) jurisdiction for almost 10 years. The chapter also examines the reconfirmation of an energy-only framework adopted in Electric Reliability Council of Texas (ERCOT), in part because of the concern that the evolving capacity framework in other US markets was not consistent with ERCOT's deregulated wholesale and retail markets.

In the economic literature, capacity markets are often viewed as a fix to the "missing money" problem where generators are not making enough revenue to recover their

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T. W:30PC open recto 1 Color 49 Lines

investment or attract new needed investment.[1] There is no evidence, however, that capacity mechanisms actually promote investment in new capacity. Furthermore, some critics claim that such markets impose an extra and unnecessary cost on customers with no proven benefits. Moreover, Moran and Skinner in Chapter 11 present a strong case that the Australian electricity market has enjoyed ample private investment in new generation for the past decade using an "energy-only" resource adequacy mechanism.

The newly proposed capacity mechanisms at the Pennsylvania-New Jersey Maryland (PJM) Interconnection and the Independent System Operator in New England (ISO-NE) extend the lead time and duration time of procuring additional resources. While these capacity mechanisms go a long way toward encouraging participation by new entrants, nonetheless, they still reinforce the old central planning paradigm that relies on multi-year constrained optimization models rather than on competitive energy markets to meet resource adequacy needs.

This chapter shows that the slow and incomplete transformation of wholesale electricity markets have limited the options that policymakers have had available in creating sustainable resource adequacy mechanisms in the United States. Capacity mechanisms in the United States have evolved in part to overcome shortcomings in transmission construction, lack of adequate demand response programs, and policies that restrict the ability of new generation to interconnect with the transmission grid. The authors contend that these shortcomings are a direct consequence of the fragmented oversight of electricity markets in the United States.

In markets that have completed that transformation into a commodities market, ding Australia, Alberta, ERCOT, and New Zealand, an energy-only mechanism has been implemented and appears to be functioning with few problems. In an energy-only resource adequacy mechanism, inframarginal energy revenues and scarcity rents derived from occasional high spot market prices are relied upon to address resource adequacy concerns.

This chapter also addresses what many policymakers consider the most contentious and troublesome issues of a sustainable energy-only approach: distinguishing between scarcity pricing and high prices caused by the exercise of market power, increasing market-based demand response, and developing a backstop mechanism that limits the transfer of wealth from load to generators in years when the reserve margin is tight. In light of the California market meltdown in 2000–01, state and federal policymakers have been reluctant to consider allowing market participants to make energy offers into real-time energy markets that are high enough to allow an energy-only resource adequacy mechanism to be successful.

The chapter ends with a proposal to use a bilateral contracting requirement for load serving entities (LSEs) that includes an energy market-based capacity product ("call option") as a backstop mechanism to transition to a sustainable energy-only resource adequacy mechanism. The proposal requires, however, that the underlying wholesale electricity market complete the transformation into a commodities market that could support such an approach.

This chapter is divided into five sections. Section 9.2 discusses concerns regarding resource adequacy. Section 9.3 presents the evolution of installed capacity markets in the US. Section 9.4 proposes market alternatives to capacity mechanisms. Section 9.5 discusses

[1] The term "missing money" was popularized by Dr Roy Shanker. See *Roy J. Shanker (2003) Comments of on Standard Market Design, Resource Adequacy Requirement*. FERC Docket No. RM01-12-000, January 10. Also see Cramton and Stoft (2006).

energy-only resource adequacy mechanisms with a focus on the experience in ERCOT. Section 9.6 presents a potential transitional mechanism based on energy-only markets with contracting obligations. Section 9.7 presents conclusions.

9.2. Concerns Regarding Resource Adequacy

Reliability of electric service has been a priority of the regulators since the inception of electric industry regulation. The regulatory approach evolved in an era when the pace of technological innovation was slower than today, and when economies of scale in the industry supported a structure of natural monopolies with vertically integrated utilities and captive load, as described in Chapter 1. In the face of rapidly changing technology and increasing opportunities for choosing wholesale suppliers – or in some cases, competitive retailers – there has been movement across the globe to unbundle, either structurally or functionally, regulated utilities.

Furthermore, through the introduction of formal wholesale and retail electricity markets, market forces were allowed to determine the mix and quantity of resources to serve end-use customers. An unintended consequence has been the potential for LSEs to lean too much on the spot markets to meet their electricity demand rather than procure sufficient resources through bilateral contracts. Insufficient bilateral contracting by LSEs can lead to a cumulative, system-wide shortfall of resources during times of peak electricity use.

Within this paradigm, regulators need to address the maintenance of a prudent reserve margin that meets the reliability needs of newly restructured markets while avoiding imposition of unnecessary expenses on customers and without excessively interfering with customer choice. In the absence of a mandatory reserve requirement to serve end-use customers, regulators need to provide incentives to ensure that sufficient and appropriate mix of resources are developed in the right locations in the grid to maintain the reliability of the grid at a competitive cost over time.

The following subsections describe historical approaches under regulatory regimes, identify the problem of "missing money" and lack of adequate incentives to attract investment, highlight perspectives on resource adequacy mechanisms, and describe the prevailing top-down resource adequacy mechanisms adopted in the United States and other countries.

9.2.1. Historical approaches under regulatory regimes

Prior to the introduction of wholesale and retail electricity markets, regulators required utilities to maintain a target reserve margin,[2] ranging somewhere between 15% and 25%, to ensure a desired level of reliability and resource adequacy. Regulators ensured that integrated utilities would serve retail load and meet prevailing reliability standards at just and reasonable prices. This approach required a mix of resources designed to serve the fluctuating demand at least cost. New capacity expansion plans were based on forecasted load growth and reserve margin requirements so as to reliably meet future needs. Regulators set the price of electricity service so as to provide the utility with an opportunity

---

[2] Reserve margin and capacity margin are measures used in the electricity industry to determine the percent of additional capacity above expected demand. Reserve margin measures excess capacity as a percent of expected demand while capacity margin shows excess capacity as a percent of existing capacity.

Els UK Job: CDI ch09-i047172 16-11-2007 4:15p.m. Page:330 Trim:165×240MM Float:Top/Bot TS: Integra, India

for cost recovery and reasonable return on its investment.[3] In some jurisdictions, up to the middle of the 1990s, utilities could build new generating facilities only if the new additions were approved in their Integrated Resource Plans (IRPs), which required utilities to review alternatives to constructing generating facilities including purchases from non-utility qualifying facilities, demand-side management, renewable resources, etc.[4]

Sending timely and accurate price signals to the majority of customers was not a key priority. However, in maintaining a healthy reserve margin, regulators used concerns over reliability and cost as justification for their policy of special tariffs for interruptible load, which were primarily targeted at large commercial and industrial customers. The regulatory approach was appropriate in an era when regulators were primarily concerned with controlling monopoly utility profits, technological innovation was slow, and economies of scale in generation and transmission justified viewing the electric power industry as a natural monopoly. However, technological innovation leading to improvements in the efficiency and cost of relatively small power plants and the promise of further technological innovation in metering control, generation, and transmission technologies prompted market reforms that allow LSEs to take advantage of the wider mix of resources available to meet their future needs. The complexity of choices, including price options, now feasible at reasonable cost, can reflect the diversity of consumer tastes, preferences, and the willingness of producers and consumers of electricity to accept more risk in exchange for proper rewards.

Unfortunately, such expanded choices in wholesale and retail markets open the door for free riding by LSEs when it comes to risk management and public goods such as service reliability. The situation is potentially more complicated in most of the deregulated retail electricity markets, where regulators may not require that competitive retailers procure resources that provide deliverable energy to their loads in a multi-year time frame.[5] The glut of generation in some of these newly restructured markets may have also reinforced the practice of relying on spot market procurements and avoiding the burden of long-term contracts in serving load that might switch suppliers in the future.

Within this paradigm, regulators need to maintain a prudent reserve margin that meets the reliability needs of newly restructured markets while not imposing unnecessary expenses on customers that may also unnecessarily interfere with customer choice. At present, most regulators do not enforce a reliability requirement on LSEs similar to what the integrated utilities in the old regulated world had to meet.

In the absence of a mandatory reserve requirement to serve end-use customers and to maximize profits by reducing unnecessary expenses or increasing market share, some LSEs may have strong incentives to procure a convenient mix of resources, rather than a mix of resources that will allow a wholesale market to maintain reliable service at reasonable cost. This desire to maximize short-term profits also may lead to underinvestment in "iron

[3] An exception to this uniformity was an interruptible tariff for industrial load.

[4] For instance, the Public Utility Commission of Texas as well as many other state regulators required regulated utilities to rely on competitive solicitation to procure additional resources to meet their increasing growth in demand for electricity. Competitive solicitations would include conventional generation resources, new technologies, such as renewable resources, and demand-side management programs.

[5] When retail markets are still relatively young, competitive retailers are uncertain about their market shares and might be at a competitive disadvantage if they procure firm resources 2 or 3 years in the future while their competitors might not do so and while loads have the financial incentive to buy from retailers with the lowest prices.

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T.W:30PC open recto 1 Color 49 Lines

on the ground" at certain times in the electricity business cycle. In the absence of any mandatory requirement to maintain a healthy reserve margin, some LSEs have the ability to lean heavily on the spot market to the extent that their credit limits permit rather than securing resources through long-term contracts.[6] Furthermore, since widespread forward contracting has the effect of suppressing real-time prices, such excess reliance on the spot market may be quite tempting. Such free ridership arises from the fact that current market rules and the installed metering and control technologies do not allow the exclusion of those who do not pay for reserves from enjoying the benefits of reserves.[7]

9.2.2. Problem of missing money

Most wholesale electricity markets have some form of price or offer caps and other forms of market mitigation measures to address price spikes or potential market failure resulting from excessive concentration of generation ownership and inelastic demand for electricity (Table 9.1). Unfortunately, mitigation designed to protect the public against market power abuse will also often suppress legitimate scarcity rents and inframarginal profits. In a competitive unrestricted market, scarcity rent is the difference between the value to consumers of the most valuable MWh that cannot be supplied due to limited generation capacity (i.e., the marginal demand-side offer accepted) and the marginal cost of the most expensive MWh served. In a long-run equilibrium, if we allow generators to collect such scarcity rents by letting demand-side bids set the clearing prices in times of shortage, then the scarcity rents should be exactly what are required to cover the amortized fixed cost of the marginal generating unit. Furthermore, if the technology mix is optimal, i.e., least total cost, then the combination of scarcity rents[8] and inframarginal profits,[9] which amount to the difference between the market clearing prices and the respective marginal costs, will exactly cover the amortized fixed cost of each generation technology in the capacity mix (not just the peaking units) in the long run (Oren, 2005a).

From a commercial point of view, scarcity rents and inframarginal profits are the margins needed above short-run marginal costs that allow

- owners of existing generation resources to recover their fixed costs of refurbishing existing plants to keep them operating and provide a fair market return on their capital (i.e., the market equivalent of regulated rate of return);
- potential developers of new resources to recover their cost of investment.

---

[6] Other reasons, including rate of new entry and rapidly shifting market share in the newly restructured retail market, low price or offer caps to mitigate prices, significant fluctuations in fuel prices, as well as uncertainty over future wholesale market design, also contribute to retailers' decisions not chosen to enter into many long-term bilateral contracts.

[7] The US, Australia, Canada, and other countries are currently seeing the spread of advanced meters capable of providing interval metering data, but this technology will not be fully deployed for a number of years.

[8] In a competitive unrestricted market, scarcity rent is the difference between the value to consumers of the most expensive MW that cannot be supplied due to limited generation capacity (i.e., the marginal demand side offer accepted) and the marginal cost of the most expensive MW served.

[9] Generators that are not on the margin and whose capacity cost is typically lower than that of peaking units on the margin recover their amortized fixed capacity cost from inframarginal profits (i.e., the difference between the MCPE and their marginal cost) plus the scarcity rents. This payment above marginal cost will both insure generation capacity adequacy and an optimal technology mix of generation.

Table 9.1. Price or offer caps in various electricity markets

Market	Price or offer cap
Alberta	$C1000
Australia	$AUS 10 000
California ISO (CAISO)	$400
ERCOT (2006)	$1000
ERCOT (2007)	$1500
ERCOT (2008)	$2250
ERCOT (2009)	$3000
France	No Cap
ISO-New England	$1000
Italy	No Cap
Japan	No Cap
Midwest ISO	$1000
New York ISO (NYISO)	$1000
Netherlands	No Cap
New Zealand	No Cap
NordPool	No Cap
Pennsylvania–New Jersey–Maryland (PJM)	$1000
Ontario	$C2000
Philippines	62 000 Pesos
Singapore	$SGD 4500
South America (Argentina, Brazil, Chile, and Colombia)	No Cap
South Korea	No Cap
Southwest Power Pool (SPP)	$1000
Spain	No Cap
United Kingdom	No Cap

Notes:

1. The Canadian markets (Alberta and Ontario) offer caps are in Canadian dollars. A Canadian dollar is worth $US 1.05–$US 1.10.
2. An Australian dollar is worth $US 0.90–$US 0.95 at current exchange rate. The Australian market does limit the amount of money a resource can capture on a weekly basis, after which the $AUS 10 000 offer cap drops to $AUS 100.
3. ERCOT market also has a low offer cap of $500 or 50 times Houston Ship Channel gas price index from the previous day, whichever is higher. See note under Table 9.2.
4. Wholesale electricity market in France does not have any price cap. However, it has regulated retail electricity prices which are kept fairly low by the government. As the wholesale market prices are rising, it becomes very difficult for alternative suppliers to compete with the incumbents.
5. The New Zealand market has no cap, but the highest price allowed in settlement has been $NZ 10 000. At the time of this writing, NZ regulators are reviewing the need for a formal cap.
6. NordPool has no price cap in the financial forward electricity market; however, there is a system limitation to the bids (currently EUR 2000) in the physical day-ahead auction. This limit is not viewed as a regulatory price cap, as it can be changed from day to day. Furthermore, NordPool has cross-border transmission capacity constraints between the NordPool countries and on the interface to Germany with market splitting when these cross-border interfaces are congested.
7. 45–50 Philippine pesos equals $US 1.00 at current exchange rates.
8. One $SGD is worth about $US 0.69–$US 0.70 at current exchange rates.
9. At the time of this writing, MISO has filed a plan with FERC that keeps the $1000 offer cap but would raise real-time prices as high as $3500 when MISO is using reserves to clear the energy market.

Els UK Job: CDI ch09-i047172 16-11-2007 4:15p.m. Page:333 Trim:165×240MM Float:Top/Bot TS: Integra, India

In other words, scarcity rents represent the market mechanism needed to signal resource shortages and provide incentives for new investment in resources.[10] Furthermore, this mechanism allows the loads, through demand-side offers, to determine how much generation capacity is needed. Proponents of scarcity pricing argue that in the absence of sufficient demand response, prices in shortage situations should be allowed to reflect the Value of Lost Load (VoLL) which could be in the thousands of dollars, well above any existing offer cap currently in place in most restructured electricity markets.[11]

Mitigated energy prices that often suppress scarcity rents may be insufficient for generation resources to earn enough return to cover their fixed costs, a problem that has been characterized as the "missing money" problem. Other forms of market interference motivated by reliability concerns, such as out of market procurement of resources (e.g., reliability must run or RMR), deployment of operating reserves to avoid involuntary curtailments, and reliability unit commitment (RUC), will have the effect of suppressing spot energy prices and contribute to the "missing money" problem. Consequently, a number of market participants and economists believe that market designs with existing price mitigation measures (that cap prices at $1000 per MWh or less) do not allow sufficiently high energy prices to provide adequate cost recovery to existing generators and may not induce timely construction of new resources for new resources.[12] For a more detailed discussion of the "missing money" problem, see Cramton and Stoft (2005, 2006).

While in principle a liquid forward market could address the "missing money" problem by providing adequate forward support for resource adequacy, most existing restructured electricity markets show limited amount of market-based long-term contracting. Evidently, the protection provided to LSEs by the capped spot prices reduces, from their risk management perspective, the optimal quantity of forward contracts in their portfolios and caps the price they are willing to pay for such contracts. A higher offer cap would transfer more risk to LSEs and would provide incentives for them to cover a larger portion of the quantity risk through bilateral contracts and reduce their reliance on the spot market.

9.2.3. Contrasting perspectives on resource adequacy mechanisms

There are two prevailing approaches to resource adequacy. One is a "top-down" or accounting-based approach that is driven by the question of whether generators are making sufficient income from energy and ancillary services to recover their fixed costs. This perspective, adopted in a number of markets in the US and abroad, directly addresses the "missing money" problem, as elaborated by Cramton and Stoft,[13] with minimal changes to the energy market. The top-down approach has opted to solve the "missing money" problem resulting from artificial suppression of energy spot prices by supplementing generators' income through "capacity payments" or through the introduction of an artificial short-term capacity product for which the demand is set administratively and the cost allocated to the load. The income from such a product is intended to make up the missing money and thus ensure the financial integrity of generators and attract new investment if needed.

[10] For a more detailed discussion on the role of scarcity rents in resource adequacy, see Oren (2005).
[11] For example, Hogan (2006) estimates VoLL at $10 000 per MWh.
[12] See Joskow (2005); Cramton and Stoft (2005, 2006); Alberta Department of Energy (2005). pp. 26–36.
[13] See Cramton and Stoft (2005, 2006).

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T. W:30PC open recto 1 Color 49 Lines

Capacity markets are viewed as a *fix* to the missing money problem but there is no evidence that they actually

- promote investment in new capacity;
- do not impose an extra and unnecessary cost on customers.[14]

The new proposed forward capacity market (FCM) mechanisms at ISO-NE and the reliability pricing model (RPM) proposed at PJM extend the lead time and duration time of the product. In Chapter 10, Bowring discusses the RPM in detail. These reforms go a long way toward encouraging participation by new entrants, but at the same time have moved closer to the traditional centralized resource planning paradigm by being prescriptive with regard to location and even fuel mix of new generation resources. The call option features embedded in the FCM proposal also improve the economic basis of the non-performance penalties and the market reciprocity, whereby generators receiving capacity payments must forgo peak energy rents (PER). However, these mechanisms still have the general shortcomings of capacity markets which suppress incentives for demand response, self-provision of sufficient generation resources to meet peak load, and tend to hinder risk hedging by load or through competing retailers.

Capacity mechanisms have been controversial in the markets where they are being implemented. Critics have stated that they overly compensate existing inefficient generation, rely on prices that are administratively determined, put too much investment risk on LSEs, and have stifled innovation and bilateral contracting for customers.[15] For a more sympathetic perspective on capacity mechanisms, see Bowring's discussion in Chapter 10.

The second approach, used in a number of markets including ERCOT, Australia, New Zealand, and Alberta, takes the view that providing resource adequacy is primarily a challenge of managing risk in a competitive commodities market. This "bottom-up" approach starts with the theoretical construct, discussed earlier; implying that in an ideal "energy-only" market, inframarginal profits and scarcity rents will cover capacity costs and lead to an optimal mix of generation technologies. Hence, earnings from unmitigated spot market transactions and long-term bilateral contracts that provide a risk-sharing mechanism between consumers and producers should stabilize generators' income streams and induce the proper level of investment. Thus, resource adequacy can be ensured by eliminating spot market distortions and by facilitating bilateral contracting and risk management.

To the extent that

- unmitigated energy markets are infeasible,
- markets for risk are not fully functional, or
- contracting practices fail to provide adequate incentives for generation investment,

then various forms of contracting and hedging obligations can be imposed.

In addition, centralized procurement mechanisms for such hedges can be instituted to promote resource adequacy. However, since such mechanisms rely on energy as the underlying commodity, they can only succeed if artificial barriers to efficient energy spot prices that can reflect scarcity are eliminated. In other words, customer protection through price mitigation measures must be greatly restricted and replaced, if necessary, with

[14] See, e.g., PennFuture (25 October 2006) for criticisms of RPM.
[15] See ELCON (2006) for comments on PJM, ISO-NE, and NYISO capacity mechanisms from the perspective of industrial customers and PennFuture (2006) for comments focused on PJM's RPM from the perspective of environmentalists.

mandatory hedging requirements that will complement voluntary risk management and contracting practices. Unmitigated energy prices will ensure that the capacity payments to generators through voluntarily contracts and mandatory hedges resolve the "missing money" problem.

Variants of the latter approach have been implemented in several restructured wholesale electricity markets with no capacity payments but a high offer cap with limited mitigation of resource-specific offers (i.e., an energy-only resource adequacy mechanism). The experience suggests that high offer caps can indeed provide a strong incentive for generation resources to supply electricity service, and for market-based demand response to be available. There is evidence that high offer caps have resulted in increased voluntary bilateral contracting between buyers and sellers and demand response, which has resulted in lower average spot market prices. Apparently, LSEs faced with the prospect of paying thousands of dollars per MWh on spot market procurement during shortage periods have been compelled to maintain forward supply contracts in order to avoid such outcomes. For instance, a number of retailers in the Australian market hedge their positions with option contracts on peaking generation that require the peakers to make energy offers into the real-time market on behalf of the retailer.[16] Similarly, in the ERCOT market, a retailer purchased a multi-year contract for the output of some peaking units to avoid being exposed to the spot market during summer peak.[17]

9.2.4. Current top-down resource adequacy mechanisms

As shown in Table 9.2, there is no unified approach to resource adequacy in existing wholesale electricity markets in the world. The authors find two basic approaches to resource adequacy that attempted to address the "missing money" issue differing by whether they specify quantities or prices for generation capacity.[18] Several restructured wholesale electricity markets in the US imposed a relatively low energy offer cap in the late 1990s but introduced a distinct capacity product (i.e., a capacity-and-energy resource adequacy mechanism). Payments for capacity aim to stabilize the payment stream for resource owners and energy prices for consumers, which in turn will attract investment in new capacity.

In the northeastern US markets such payment for capacity is implemented indirectly through a capacity obligation imposed on the LSEs and an adjunct market for trading and procurement of capacity credits. Recognizing that consumers are only interested in energy consumption while capacity is an artificial product, capacity markets are based on an administrative demand prescription according to a technical determination of target capacity. The administrative demand can be set at a fixed level as in the traditional ICAP markets and the current ISO-NE FCM proposal. But recently, several ISOs have adopted a demand function approach, also known as variable resource requirement (VRR), where the administrative demand for capacity is price-sensitive, thus allowing more capacity to be purchased as the procurement price declines.[19]

[16] Authors' communication with Peter Adams, Manager, Surveillance and Enforcement, Markets Branch, Australian Energy Regulator, 1 February, 2005.
[17] The retailer's name is not divulged for reasons of confidentiality.
[18] A comprehensive analysis of alternative resource adequacy mechanisms with particular focus on Europe can be found in the Ph.D. thesis of De Vries, L.J. (2004). Correljé and De Vries discuss this issue in Chapter 2..
[19] Such an approach is used by the NYISO and is part of the PJM Reliability Pricing Model (RPM).

Table 9.2. Some examples of restructured electricity markets with and without capacity payments

Level of offer/Price cap	No price mitigation	Capacity market	Direct capacity payments	No capacity payments
In some cases, cost-based Energy offers			South America, Spain, Italy, South Korea	
Low offer caps (up to $1000)	Alberta, ERCOT (2006)	ISO-NE, NYISO, PJM		Alberta, CAISO, ERCOT (2006), MISO, SPP
Medium offer caps ($1000 to $3000)	Ontario, Philippines, Singapore		Ontario	ERCOT (2009), Philippines, Singapore
No or high offer caps (above $3000)	Australia, France, Japan, New Zealand, Netherlands, NordPool			Australia, France, Japan, New Zealand, Netherlands, NordPool, United Kingdom

Notes:

1. South American markets include Argentina, Brazil, Chile, and Colombia.
2. The Australian market does limit the amount of money a resource can capture on a weekly basis, after which the $AUS 10 000 offer cap drops to $AUS 100.
3. In CAISO, Load Serving Entities (LSEs) are subject to a forward bilateral contracting obligation.
4. The ERCOT market does limit the amount of money a resource can capture to $175 000 per MW of capacity, after which the $3 000 offer cap drops to low offer cap of $500 or 50 times Houston Ship Channel gas price index from the previous day, whichever is higher.
5. At the time of this writing, MISO has filed a plan with FERC that keeps the $1000 offer cap but would raise real-time prices as high as $3500 when MISO is using reserves to clear the energy market.
6. In South Korea, capacity payment is determined annually by the Generation Cost Evaluation Committee considering the long-term marginal fixed cost of the generators, which varies by fuel type.

The alternative capacity remuneration mechanism is direct capacity payment originating from the theory of socially optimal peak load pricing in a regulated monopoly dating back to Boiteaux (1960). Under that theory, social welfare is maximized when energy is priced at marginal cost. However, since such pricing cannot recover fixed costs, a second-best approach with minimum social welfare degradation is to supplement marginal cost pricing with a capacity payment that equals to the amortized fixed cost of the marginal technology and recover these added costs as a demand charge from consumption during peak periods.

Capacity payments are popular in several South American countries such as Argentina, Chile, Peru, and Colombia, in European countries such as Spain and Italy, and in South Korea where generators receive direct payments for capacity from the system operator, which are uplifted to customers on a prorated basis. Capacity payments are sometimes differentiated according to generation technologies (e.g., South Korea has a two-track payment for base-load capacity and peaking capacity) and may be commingled with stranded cost recovery (e.g., Spain). Often such payments are coupled with a cost-based offer requirement in the energy market (e.g., South Korea and Peru).

In Argentina and Peru, only capacity that is selected to produce energy or provide reserves will receive capacity payments. This practice has resulted in Argentina of energy offers below marginal cost. Such "underbidding" is prohibited in Peru where the regulator determines the marginal cost of each generation technology except for suppliers with natural gas plants, who can declare their marginal cost annually. In South Korea there is a two-tier system for base-load and peaking generators. All generators are required to submit regulated marginal cost-based offers for energy. Generators are classified into two tiers, which receive different capacity payments, and their energy offers are cleared separately for base-load and for peaking load, resulting in two different market clearing prices for the generators. The load is charged an average price that recovers the energy and capacity payments to the generators.

In general, capacity payment mechanisms are an effective means to stabilize the income of *incumbent* generators, but there is no clear evidence that such payments encourage investment in *new* generation. Generators receiving capacity payments have no obligation to use that income for new investment or for improvement in their facilities. Furthermore, since capacity payments have the effect of suppressing marginal cost and resulting energy spot prices to consumers who cannot opt out of paying the capacity charges, such mechanisms undermine potential demand response. In Chile, one of the first countries to liberalize their electricity industry and to institutionalize a capacity payment scheme, recognition that this scheme has not produced the desired outcomes in terms of providing incentives for investment has led to recent re-assessment of their resource adequacy policy and new proposals along the line of call option obligations.[20]

9.3. Evolution of Installed Capacity Markets in the US

In the early days of deregulated wholesale markets in the eastern US, capacity owners were under a must-offer obligation to offer their capacity into the market for a given price ($/MW). Such arrangements evolved from resource sharing agreements that existed in long-standing power pools such as PJM and NEPOOL, now ISO-NE. In these agreements utilities had an obligation to maintain sufficient capacity to cover their peak loads and were penalized for capacity shortfalls. The establishment of a market for short- term capacity credits, later referred to as an Installed Capacity Market, or ICAP, was intended to enable utilities with excess capacity to be compensated by utilities with capacity shortfalls when the overall capacity available in the system was sufficient to meet total peak load and reserves requirements.

Such market, however, had several shortcomings. Because of their short duration, only existing "steel in the ground" could participate, resulting in the so-called "bipolar pricing," where the price was either zero when the total available capacity exceeded the total requirement or as high as the shortfall penalty when there was a capacity shortage in the system. In addition, while the total capacity might have met the aggregate capacity requirement in the system, transmission constraints often made available capacity non-deliverable to load pockets where it was needed, resulting in local reliability problems. Because of the short duration of ICAP products, ICAP markets were also susceptible to exercise of market power due to non-contestability by new entrants. At the same time, energy markets have not been able to provide correct price signals and sufficient revenue to provide incentives for new entrants due to the lack of scarcity pricing

[20] See Barroso et al. (2006).

that would reflect deployment of operating reserves to prevent energy shortfalls. These deficiencies have been widely recognized and, over the last several years, significant efforts have been devoted to improving the capacity markets in the US, as described later.

In this section, the authors first review earlier versions of ICAP payments established in the northeastern US markets, particularly in PJM. This review will also include minor modifications that were made to these mechanisms to address some of the earlier problems. Next, we will briefly discuss new and improved version of ICAP markets that were developed through further co-operation between stakeholders and state and federal regulators to improve the performance of these capacity market mechanisms. The mechanisms discussed later include the NYISO's extended product duration and sloped demand function, the PJM's 3 year forward-looking capacity market (Reliability Pricing Model or RPM), and the ISO-NE's 3 year forward-looking capacity market (FCM).

The following subsections describe the early ICAP markets and their minor refinements, followed by a discussion of recent steps taken in some markets to introduce new and improved capacity mechanisms.

9.3.1. Early ICAP markets and their refinements

The PJM installed capacity market was the first of its type to become operational in 1999. Capacity owners were under obligation to offer their capacity into the market for a given price ($/MW). There was no distinction between old and new capacity, and capacity owners would be penalized if they could not meet certain performance standards. The ICAP payments did not take into account the value of the location of the resource or the operational characteristics of the unit. For a more detailed discussion of capacity markets, see Bowring's discussion on capacity markets in Chapter 10.

This approach proved inadequate for assuring resource adequacy, as the US Federal Energy Regulatory Commission (FERC) noted in its 2004 Staff report, stating that

> Much of the country has no obvious market mechanism to signal the need for new building in advance of shortages. The success of capacity markets in addressing the issue is not yet proven.[21]

9.3.1.1. Shortcomings of the original PJM ICAP

The original PJM ICAP market had the following four shortcomings:

1. Bipolar nature
2. Deliverability problems
3. Lack of contestability
4. Lack of scarcity pricing

Bipolar nature of ICAP clearing prices The first incarnation of an ICAP market has been the most contentious aspect of ICAP. This version of ICAP was a monthly product. From the generators' viewpoint, a one-month installed capacity obligation does not provide sufficient certainty of cash flow to attract capital for new generation investment that would be built for years in the future. Furthermore, the short look-ahead period of the ICAP obligations in the early design of the PJM market left no room for participation in the

[21] Federal Energy Regulatory Commission (2005). p. 65.

ICAP market by the newly planned capacity due to the length of time that it takes for such capacity to be built.

Consequently, the ICAP markets have been limited to existing capacity, and the result of that was bipolar pricing: the value of the ICAP product was either zero (during times when the reserve margin was healthy) or infinity (during times when the reserve margin was thin).[22] The high price volatility and eventual collapse of the initial daily ICAP market at PJM led to the development of a more sustainable monthly capacity market and to a proposal for a seasonal capacity obligation.[23]

Similar moves toward capacity products with longer durations have been implemented at ISO-NE and NYISO. The PJM ICAP market currently operates with multiple maturities: 12 months, 6 months, and 1 month, as well as a daily market. The NYISO operates markets for capacity obligations of 6 months, multiple months up to 6 months, and 1 month.

Deliverability problems of ICAP products Inability to build transmission quickly enough to accommodate new generation investment and concentration of new capacity in locations that are remote from load pockets where generation capacity was needed results in deliverability problem in meeting expected electricity demand. For example, in ISO-NE new quick response units were built in Maine near the gas sources but energy from these resources could not be delivered to Connecticut load centers because of transmission constraints.

Lack of contestability with new generation As indicated earlier, the short-term nature of the capacity product made it impossible for new generation capacity to participate in the ICAP market so that incumbent generation with steel in the ground could exercise market power in shortage situations. The reforms extending the duration of the ICAP obligation did not go far enough in terms of creating capacity obligations that could enable responses by the newly planned investment when ICAP prices increased due to capacity shortages.

Lack of scarcity pricing In cases of energy shortage the system operators tend to draw energy from operation reserves. Using reserves augments the balancing energy offers curves and results in suppressing scarcity rents that would have resulted if the shortage was mitigated through demand-side offers. To avoid such suppression of price signals, energy spot prices must reflect the cost of reduced reliability resulting from the depletion of operating reserves, but such scarcity pricing was not introduced in the early market designs. Similar price suppression results from deployment of "reliability must run" units by the system operators in order to assure local reliability when markets fail to provide needed local resources.

Over time, it has become apparent in US electricity markets that setting clearing prices based on tightly mitigated resource-specific offers do not send appropriate scarcity pricing signals in real-time when the system operator needs to deploy operating reserves. Nodal markets in the eastern Interconnect of the US are instituting various forms of scarcity pricing to send a more appropriate price signal in real-time, allowing prices to rise more frequently to the $1000 offer cap in each market, improving short-term reliability of the grid

[22] While the value of the product would be infinity if the market price were unconstrained, in reality the price of the monthly ICAP product cleared at the cap set by PJM.
[23] Federal Energy Regulatory Commission (2002).

through market-based mechanisms rather than a command-and-control approach.[24] These markets are moving toward or have implemented real-time co-optimization of energy and ancillary services. In addition, one or more of the markets use an administrative pricing mechanism that adjusts real-time energy offers to reflect whether the real-time energy offer stack is being or has been depleted.[25] Additional cost recovery for generation has been a by-product of this change rather than a driving force behind it.

9.3.1.2. Centralized Resource Adequacy Market (CRAM)

The deficiencies in existing ICAP markets and seams problems arising from incompatibilities in these markets prompted the commissioning of a study to develop a proposal for a single integrated capacity market for the ISO-NE, the NYISO, and PJM. The NERA (2003) report recognized the need for a long-term forward-looking capacity obligation to ensure adequate investments in generation and proposed a central resource adequacy market (CRAM) for these three markets. The report proposed that LSEs be subject to a capacity obligation to assure that sufficient capacity is in place with sufficient lead time for planning and construction, as suggested by FERC in its Notice of Proposed Rule Making that proposed a Standard Market Design. The ISOs would act as central buyers of capacity and make forward commitments to buy capacity. These commitments would be supported financially by uplifted charges to all LSEs during the capacity supply period. According to the CRAM proposal:

> the ISO would determine the resource need in advance of the planning period, would hold a central procurement through an auction, would pay the auction price to all resources provided during the period and would recover the cost from load during the planning period. The difficulties arising from uncertainty with respect to load obligations several years in the future would be eliminated and all LSEs would face a common charge for resource adequacy that would be passed on to consumers and would be competitively neutral at the retail level. Consumers will receive the benefits of adequacy and pay the cost of adequacy.

A key aspect of the proposed scheme was that

> The planning horizon must be sufficiently long to enable the CRAM to be a deciding factor in the decision to construct. …Only when the pool of competitors is expected to include entrants can market power concerns be adequately addressed. Practically, this means that a three-year planning horizon is the minimum. [26]

The proposal recommended that the commitment period should be from 1 to 3 years with preference for longer durations in order to reduce generators' uncertainty about revenues – which is expected to result in lower risk premiums in their costs of capital. The proposal suggested that all required capacity be procured or under contract at all times, arguing that sequential auctions for progressive procurement would be unreliable for determining

---

[24] McNamara, Ron (2006). Midwest ISO, resource adequacy in Midwest energy markets. Presentation at the Organization of MISO States, May 8–9, 2006. Available at http://misostates.org/R1-%20Ron%20McNamara%20-%20MISO.pdf

[25] Starting in 2005, the NYISO began using an operational reserves demand curve when it is short of operational reserves. ISO-NE, PJM, and MISO have reviewed the issue and are in the process of implementing a similar approach.

[26] NERA Economic Consulting (2003). *Central Resource Adequacy Markets for PJM, NY-ISO and NE-ISO, Final Report*. February.

prices. The CRAM proposal was shelved due to disagreements among interested parties in the three North Eastern ISOs. However, some of the elements of the CRAM proposal have resurfaced recently and influenced capacity market designs adopted by ISO-NE and by PJM.

9.3.1.3. ISO-NE's proposed locational ICAP market (LICAP)

The New England ICAP market has experienced bipolar pricing of its capacity deficiency auction like other ICAP markets. Prices were fluctuating between zero and the capacity deficiency penalty, but prices eventually collapsed due to system-wide excess capacity, as shown in Fig. 9.1. After February 2004 ICAP prices have essentially dropped to zero.

Litvinov et al. (2004) reported that ICAP prices have been insufficient to support existing generation and new investment. Furthermore, because the capacity market in New England does not account for transmission constraints, system-wide excess capacity in the ISO-NE territory has masked local deficiency of capacity in congested areas such as Boston. Out of the total generating capacity in New England, which amounts to 32 615 MW, 41% is financially distressed due to revenue shortfalls and the owners are in various stages of bankruptcy. Figure 9.2 (reproduced from Litvinov et al., 2004) illustrates the average net revenue shortage for combined cycle (CC) and combustion turbines (CT) units under alternative estimates of the carrying costs for these units. The carrying cost includes cost of capital, plant O&M cost, and other economic parameters such as tax, inflation, and risk adjustments.[27]

The estimates of required fixed cost recovery shown in Fig. 9.2 motivated, in part, the ISO-NE's proposed solution of creating locational ICAP markets for four distinct areas: Maine, Northeast Massachusetts/Boston (NEMA), Connecticut, and the rest of the ISO footprint. Co-ordinated ICAP auctions account for three directional capacity transfer

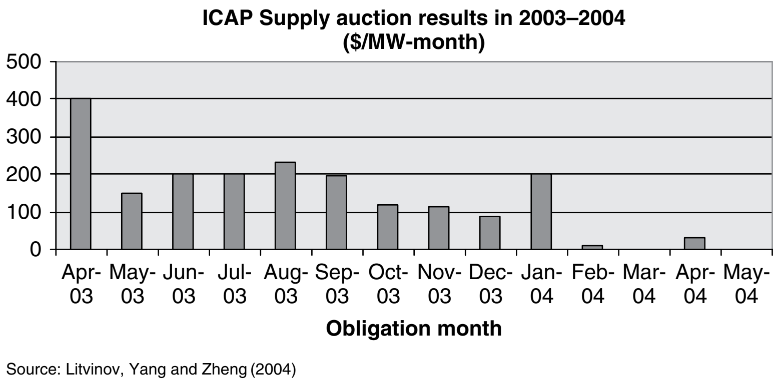
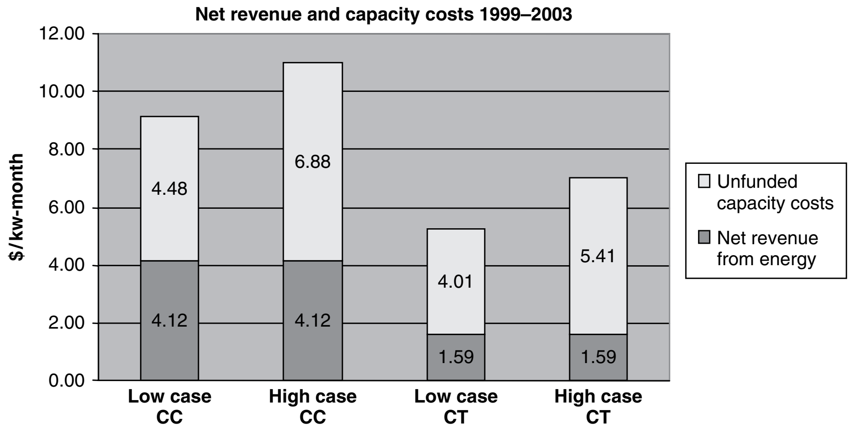


Fig. 9.1. Prices in the ISO-NE ICAP market.

[27] In contrast, according to Potomac Economics(2006). [*2005 State of the Market Report for the ERCOT Wholesale Electricity Markets*. July, pp. 49–50] wholesale electricity prices in ERCOT in 2005 greatly exceeded the thresholds for fixed cost recovery for base-load coal and nuclear power plants. Given the projections for strong economic growth and natural gas prices in ERCOT for the foreseeable future, market participants have informed ERCOT planning staff that they are considering investing tens of thousands of MWs of new coal-fired, nuclear, and wind capacity in ERCOT through 2010.

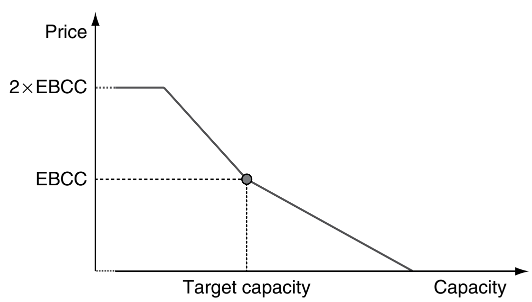


Sorce: Litvinov, Yang and Zheng (2004)

Fig. 9.2. Average net revenue shortage of combined cycle and combustion turbine generators in the ISO-NE.

constraints: Maine exports, NEMA imports, and Connecticut imports. Financial capacity transfer rights (similar to flowgate rights) would be issued to provide instruments for hedging locational ICAP (LICAP) price differences.

A demand curve for LICAP was to be used in the ISO-NE spot LICAP auction. According to Cramton and Stoft (2005), the demand curve was intended to provide a rough approximation to a capped annual energy revenue stream, including scarcity rents that would naturally decline as reserve capacity increases and scarcity rents go down. The proposed LICAP demand curve, illustrated in Fig. 9.3, was anchored at a nominal level of target capacity with a corresponding price that equals the amortized expected capacity carrying cost of a peaking unit. From that reference point, currently set to 106% of the minimum LICAP requirement, the curve extends linearly in both directions. The demand curve intersects the $0/kW per month level at some preset capacity value above the target level. Currently, that point is set to 118% of the minimum requirement. The demand curve



Note: EBCC = expected benchmark carrying cost (annualized fixed cost of frame unit)
Source: Cramtonand Stoft (2005)

Fig. 9.3. Proposed ICAP Demand Function in ISO-NE.

rises as quantity drops below the target level and is capped at twice the expected carrying cost of a peaking unit when capacity is at or below the minimum LICAP requirement.

The LICAP clearing price was to be adjusted *ex post* on an annual basis by subtracting the inframarginal energy revenues per MW per year realized by a combustion turbine (CT) used as a benchmark. The adjusted clearing price was intended to settle LICAP shortfalls or excesses among LSEs and to determine payments to generators and dispatchable load resources. Payments were to be prorated based on availability during a predetermined set of days when generation capacity is scarce. The locational feature of LICAP and the prorating of payments based on availability on certain days were intended to add intrinsic value to an otherwise artificial product whose demand is derived from an administrative requirement. Likewise, although artificially determined, the downward sloping demand function provided an effective means of eliminating the binary nature of ICAP prices, the result from a vertical demand function, and it eliminates much of the incentive to withhold capacity.

Eventually the LICAP proposal was scrapped due to strong opposition from different advocacy groups, a strong lobby in the US Congress, and governors in five out of six states in the ISO-NE jurisdiction.

9.3.2. New and improved capacity mechanisms

Given some of the shortcomings mentioned above regarding the earlier ICAP markets in the Northeastern US, market operators and stakeholders along with scholars identified new approaches to improve the operation of capacity mechanisms. In fact, new and improved versions of previous capacity markets were introduced in all three ISOs (PJM, NYISO, and ISO-NE), and FERC eventually approved their implementations. The improved mechanisms are briefly described in the following subsections.

9.3.2.1. PJM's reliability pricing model

To improve its capacity mechanism, PJM recently reached a stakeholders consensus in finalizing its Reliability Pricing Model (RPM) with a variable resource requirement (i.e., sloped demand function), which was filed with FERC on August 31, 2005. RPM is based on an integrated resource planning model that looks to4 years into the future to determine generation resource needs in terms of location and fuel mix. Under the originally proposed scheme, the needed generation capacity to maintain adequate reliability is procured through a central auction on a 4 year forward basis, which enables participation by existing generator and new investors. RPM encompasses existing and planned transmission, generation, and demand-side response, as well as incorporating locational pricing in its forward auction. Bowring discusses this mechanism in more detail in Chapter 10.

The RPM auction features an administratively determined downward sloping demand curve that allows the procured quantity to vary with price. This scheme is also known as variable resource requirement (VRR). The use of an administrative demand function has been rationalized on the ground that it reduces volatility of the capacity payment to generators and thus encourages more investment in generation capacity, resulting in increased social welfare.[28] However, this argument is debatable since it is based on the assumption that generation firms are risk-averse, while in calculating the social benefits

[28] See testimony by Ben Hobbs in FERC Docket No EL05-148-000 and ER05-1410-000, Initial Order on Reliability Pricing Model.

society as a whole and consumers in particular are assumed to be risk-neutral. The intent of RPM was to be compatible with areas having retail choice as well as areas with traditional regulation.

After the initial approval by FERC[29] on April 20, 2006, the parties entered into 4 months of settlement discussion (ordered by a FERC Administrative Law Judge) that resulted in only slight modification to the original proposal. Specifically, the contracting lead time was reduced from 4 to 3 years and the demand function was shifted down so that it is capped at 1.5 times the cost of new entry (CONE) which is estimated at about $65000/MW per year. The curve drops linearly from 1.5 × CONE at 98% of target capacity to 0.2 × CONE at 105% and then vertically down to zero. The RPM establishes locational capacity requirements, allows for demand response and transmission participation, has explicit market mitigation rules, and allows opt-out alternatives for LSEs that do not want to participate. Features such as seasonal pricing of capacity, operational price adders, and load following requirements for portion of the capacity obligation, which were in the initial proposal were eliminated, while some other minor features have been added.

9.3.2.2. NYISO's Demand curve model

In an attempt to reduce such volatility of ICAP prices that had been at the NYISO and other ISOs, the NYISO was the first to introduce a variable resource requirement, also known as an ICAP demand curve, in the New York capacity market.[30] The demand curve model was developed through the stakeholder process and came into effect in May 2003. Prior to introducing VRR, the NYISO ran a semi-annual auction for 6 month capacity products and a monthly capacity auction for monthly capacity products for the remainder of the 6 month capability period, as well as a centralized deficiency auction prior to each month.

Each LSE had to provide contracts to demonstrate to the NYISO that it was covering its capacity requirement for the ensuing month. Any shortfalls were covered through the centralized deficiency auction in which the NYISO bid for all the deficient capacity at a price equal to the deficiency penalty imposed on LSEs for each MW-month of capacity deficiency. LSEs exceeding their capacity obligation could offer their excess in the auction. The deficiency auction represented a "vertical demand function" where the ISO demanded a fixed quantity of capacity, and resource providers and LSEs with spare capacity offered supply schedules against it. The experience has been that prices in that auction were either at the deficiency price or close to zero.

Under the VRR arrangement, the 6 month and monthly ICAP auctions continue to operate as double auctions in which LSEs bid for supply and resource providers offer supply. The deficiency auction, however, has been replaced with the VRR, which represents a downward sloping demand curve capped at the deficiency penalty. The downward sloping segment of the demand function is linear and determined by two points. The first "reference point" is defined by the minimum capacity requirement and a price that equals some percentage of the estimated cost of a CT. The second point is set at the level of capacity at which the capacity value is nil. The parameters of the ICAP demand curve

[29] FERC Docket No EL05-148-000 and ER05-1410-000.

[30] New York Department of Public Service (2003a); New York Department of Public Service (2003b); New York Independent System Operator Inc. (NYISO) (2004a); and New York Independent System Operator Inc. (2004b).

vary by location (specifically differentiating New York City (NYC) and Long Island from the rest of the state) and are subject to adjustment. According to the NYISO Tariff 2004b (section article 5, section 5.14.1(b)), the reference point for NYC as of May 1, 2004 was set at $151.14/MW per day for 100% of the minimum local ICAP requirement, and from there the curve declined linearly to $0 at 118% of the minimum ICAP requirement.

In the NYISO report to FERC[31] it is stated that the ICAP demand curve has achieved the goal of stabilizing ICAP spot prices in the deficiency auction. Furthermore, purchased quantities in the deficiency auction have increased, while clearing prices have decreased. The deficiency auction has also provided a price floor for the 6 month and monthly capacity markets. The VRR seems to function well and mitigates incentives for withholding capacity by rewarding available capacity in excess of the minimum requirement and by recognizing that such extra capacity has value in enhancing reliability and moderating energy and ancillary service prices.

9.3.2.3. ISO-NE's forward capacity market

Likewise, ISO-NE went through a consensus-building process with stakeholders and state regulatory authorities to finalize a settlement agreement to address its capacity market. In April 2006 a FERC Administrative Law Judge approved this settlement agreement that outlined a forward capacity market for ISO-NE that will replace the previous LICAP design. The new agreement envisions a 3 year forward-looking capacity market (FCM) where the capacity product emulates some key features of physical call options and the procurement is done through a descending clock auction with a vertical demand curve.

The FCM has integrated some key features that are similar to the call option approach described earlier and some element of the CRAM approach. Capacity is procured 3 years forward for duration of 1 year for incumbent generators and up to 5 years for new generation through an annual competitive descending clock auction. The procurement is zonal, based on local reliability needs, load forecasts, and forecasted transmission availability and the capacity payment could vary by location. The procured capacity contracts must be backed by qualified existing or planned physical generation resources or by demand-side resources.

The procured capacity contracts entail an obligation to offer energy in the spot energy market at a strike price that reflects the marginal energy cost of a generic peaking gas turbine unit. The strike price is enforced by deducting from the capacity payments the peak energy rents (PER), which represents the excess revenues of a generic peaking unit computed for the annual duration of the contract[32]. The capacity payments for multi-year contract obtained by new generators are based on the auction price of the first year and indexed for inflation in subsequent years. Payments for the capacity and cost recovery from the load occur at the time of performance (starting 3 years after the procurement).

Periodic and seasonal reconfiguration double auctions enable parties to adjust their positions by committing additional capacity or withdrawing (delisting) committed capacity. Such delisting can take place up to 4 months prior to start of the commitment period.

[31] FERC (2004). *Report on Implementation on ICAP Demand Curve*. New York Independent System Operator, Inc., Docket No. ER03-647-000.

[32] The PER deduction are capped at the auction-based capacity payment so that the net capacity payment cannot be negative.

The seasonal reconfiguration auction also allows changes in procured capacity so as to reflect changes in load forecasts and for trading of obligations among market participants. The traded tender in the reconfiguration auction is full-year commitment, unlike the traditional ICAP markets (e.g., NYISO) where the traded tender has been monthly or even with shorter-term capacity obligations. Non-performance is subject to penalties in the form of reduced capacity payments. In that respect the FCM contract differs from call options since the non-performance penalties are not based on actual liquidation damages reflecting the difference between the spot energy cost and the strike price. The FCM contract is designed to prevent non-performance penalties in excess of the net capacity payment (after PER adjustment) received by the non-performing resource, which would be possible under market-based penalties, reflecting actual liquidation damages. A transition mechanisms employing straight capacity payments will be implemented to assure continuity of capacity-based income to incumbent resources.

9.4. Energy Market Alternatives to Capacity Mechanisms

The evolution of capacity markets in US East Coast nodal markets has been a gradual realization through experience that undifferentiated capacity and its related energy are insufficient to meet all the requirements of operating a wholesale electricity market efficiently and reliably. Furthermore, experiments with retail markets in the eastern Interconnect have been mostly unsuccessful to date, with at least one state, Virginia, moving to re-regulation while others have been in turmoil. The notion that consumer choice will determine the level of reliability and the resulting need for generation capacity seems untenable at this time in many jurisdictions. As a result, regulators have favored centralized planning based on engineering consideration and constrained optimization methods in determining the capacity needs while limiting the role of market mechanisms such as auctions, request for proposals (RFPs), and centralized capacity markets to procurement of these predetermined quantities.[33]

The evolution of capacity mechanisms to resource procurement mechanisms, such as seen with RPM, raises the question, which we address in the following subsections: “Why not let energy market forces guide resource adequacy decisions?”

The following subsections briefly describe energy market-based resource adequacy mechanisms with no direct capacity remuneration and assess the potential benefits of such an approach.

9.4.1. Energy market-based resource adequacy mechanisms

It has been argued that capacity mechanisms will always be a required part of a deregulated electricity market because regulators, policymakers, and the public are reluctant to accept occasional high and volatile energy prices that would be required under an energy-only approach. The concern about high prices in spot markets is clearly a potential political impediment to the implementation of an energy-only resource adequacy mechanism. The

[33] Using a multi-year constrained optimization as part of a resource adequacy mechanism is theoretically sound as the duality theory shows that outcomes of a constrained optimization and a set of centralized spot markets are equivalent in the absence of market power. See Charles Rivers and Associates [Ruff, Larry] (2004). *A Transitional Non-LMP Market for California: Issues and Recommendations*, pp. 5–6.

troubled evolution of capacity mechanism in the eastern Interconnect has highlighted, however, a more fundamental issue in meeting resource adequacy in those wholesale markets, which is the halting transformation of electricity deregulation itself.

The preference or even necessity for capacity markets that feature a centralized mechanism that relies on constrained optimization over a 3 or 4 year period reflects a symptom of basic problems in electricity market infrastructure. After all, one of the drivers of electricity market reforms has been general discontent with the traditional vertically integrated local monopoly-based electric power industry whose decisions and performance incentives were encumbered by the regulatory compact. One of the primary motivations for reforms was the recognition that like in other critical infrastructure industries, market forces could lead to better investment decisions and innovation.

Approaches such as RPM and FCM appear to be taking the electricity deregulation away from the fundamental goal of electricity deregulation. The goal of electricity deregulation is the transformation of an industry dominated by regulated monopolies into a competitive electricity market that looks, as much as technically feasible, like a commodities market and can deliver to customers the benefits of economic efficiency. A competitive commodities market requires easy entry of new suppliers, good transportation networks, liquid spot and forward markets, and vibrant competition on both the wholesale and retail levels. To function reliably and efficiently, a wholesale electricity market needs the right mix of base-load, load-following, and peaking generation. The system operator needs the appropriate mix of generation and load resources to meet real-time contingencies.

The following fundamentals, some of which are not part of a number of electricity markets in the US and abroad, are necessary for electricity markets to truly become a fully fledged commodities market that can allow market forces to meet resource adequacy needs while maintaining system reliability:

- Easy interconnection of generation: Generation must be able to respond to locational market prices by quickly siting at the appropriate location in the electric grid where additional supply is needed. Difficulties in generation siting undermine the power of scarcity pricing signals and can increase the potential for market power abuse.
- Proactive investment in new transmission: Delivery of energy to all loads at competitive costs is important to create genuine competition among suppliers for all loads. Many load pockets, particularly in the United States, are associated with areas that are in non-attainment for ozone or other air quality standards, making the siting of new generation in those locations problematic. Proactively reducing transmission bottlenecks and anticipating load growth would encourage the import of competitively priced power from outside of the load pocket.
- Socialized payments by all loads of new transmission: In certain markets in the United States, regulatory authorities spend substantial time and effort deciding what portion of a proposed transmission expansion is needed for reliability (the cost of which is paid by loads) and what is needed to improve economic dispatch (the cost of which is paid by generation). This approach greatly reduces the amount and speed of transmission construction and adds risk and cost to the siting of new generation if a generator needs to pay a substantial and undetermined portion of the transmission construction costs. The experience in ERCOT within Texas, Alberta in Canada, and in Australia suggests that socializing the payment of transmission construction allows for smooth and quick entry of new generation resources in the wholesale market. This may turn to be the most economical way to improve market efficiency and result in significant benefits to consumers by further fostering wholesale market competition.

- Enhancing demand responses: Having price responsive load increases competition for generation at or near peak demand, reducing the need for or scope of *ex ante* mitigation out of concern for system-wide market power abuse while allowing for scarcity pricing. In Chapter 8, Zarnikau presents a detailed discussion on demand response.
- Bilateral forward contracting: Bilateral forward contracting can provide a myriad of energy risk management features customized to the preferences and abilities of end-use customers. Capacity markets that do not use an option approach can greatly restrict the type and size of demand response because of the rigidity of the associated regulatory process and the long lead times for procuring new resources.
- Retail competition: Competition removes the agency problem of incumbent utilities that own generation. These utilities may hinder or fail to respond to opportunities to adopt new load management technologies that would undermine the profitability of their generation holdings. Retail competition allows for a wider range of options in pricing the use of electricity by end-use customers by relying on the dynamic creativity of the market rather than a slow and unresponsive regulatory process to meet the preferences of end-use customers.
- *Regulatory co-ordination*: Energy-only markets, such as Alberta, Australia, ERCOT, and New Zealand, have a much stronger degree of regulatory co-ordination than is seen in FERC jurisdictional markets, where the US Federal government is the regulator of the wholesale market and each individual US state government is the regulator of the state's retail load. Jurisdiction of transmission planning and payment for new transmission is even more fragmented in FERC-jurisdictional markets. Because well-functioning electricity markets are fiendishly complex with numerous interrelationships, regulatory co-ordination is vital in developing and maintaining the necessary conditions for an effective energy-only resource adequacy mechanism with retail customer choice.

9.4.2. Benefits of an energy market-based approach

The benefits from using an energy market-based approach to resource adequacy are the efficiencies derived from having energy price risk managed by loads and investment risk managed by generation developers.[34] As described in the previous subsection, this optimal mix of price and investment risk has failed to materialize in most markets because the market transformation itself has been incomplete. These risks are no greater than those risks associated with other commodity and financial markets in the world today provided that the enabling technologies and institutional arrangements for demand side participation in the energy are in place. Managing these risks will encourage market participants to develop more cost-effective ways of delivering and consuming electricity through bilateral contracting while relying on spot markets for adjustments in their position in response to unforeseen load fluctuations or contingencies. The result will be a more efficient use and production of electricity that will serve the same amount of economic activity within a given competitive electricity market at reduced cost.

In the regulated world, the rates that customers paid included implicit capacity payments for generating units regardless of whether they were used continuously as base-load units

[34] The discussion in this section is based on Schubert (2005).

or only a few times a year as peaking units. Although regulators required utility assets to be used and useful, the justification for new units presumed an obligation to serve and that all customers, except interruptible customers, were entitled to firm services. By determining an identical, fixed level of reliability of electricity service for each customer, the regulators implicitly set a value of reliability that was based on statutory and policy decisions and quantified in terms of engineering criteria.

In a deregulated market, units with low-capacity factors need to earn sufficient revenues to be kept online; however, they may not earn enough money because a number of market participants do not wish to pay for a very high level of reliability or are not required to do so by a regulatory rule or market protocol. As an alternative, the market relies more on price-responsive demand and peaking shaving to reduce the number of units with low capacity factors while maintaining system reliability.

In reality, if the price of electric service rises to a certain level, some market participants may be willing to curtail their electric service voluntarily for a number of hours each year rather than pay peaking units to deliver that power during those high-priced hours. These market participants would consider the old standard of reliability inappropriate for their needs, for at least a portion of their electric service. An efficient market design would reflect the different values of reliability (interruptibility) they place on a portion of their electric service by letting market prices reflect those preferences provided that such service quality differentiation is technically feasible.

Under an energy-only resource adequacy mechanism, high offer caps could provide needed potential "headroom" for commercial and industrial loads, in the form of price-responsive loads to compete with the traditional peaking units, the old standby of the regulated era that can sit idle for 90 percent or more of the year. This competition with traditional peaking units would take place in the real-time energy and ancillary services markets in a number of ways: easing ramping constraints in the market, managing load during summer peaks, and providing services to the system operator in the form of reserves and possibly regulation.

The possibility of high prices combined with new technologies could provide incentives for smaller loads to increase their deployment of supply- and demand-side alternatives such as solar panels, energy efficiency appliances, and controllable loads.[35] Higher summertime prices, combined with real-time pricing for a variety of customers, including residential, would provide stronger incentives for installing energy-efficient air-conditioning, increasing insulation of buildings and homes, and creating demand for smart appliances like pumps for swimming pools and hot water heaters that can be set to operate when prices are low.[36] Zarnikau presents addition examples in Chapter 8.

---

[35] An example of a controllable load is an air-conditioning cycling program. When prices and demand rise to a certain level or when market operator declares an emergency condition, a device on an air-conditioner receives a signal, which turns off the air-conditioner for a certain amount of time. When a large number of air-conditioners are aggregated under one controller, the amount of demand responses can be calibrated by the number of air-conditioners cycled at one time. In this capacity, the portfolio of air-conditioners can mimic a small generator by incrementally increasing or decreasing the energy used by the portfolio and could provide ancillary services to grid operators.

[36] Time-of-use pricing, with predetermined rates reflecting higher average costs of generating electricity over peak summer hours, would be one way for residential load to be exposed to price risk. Large industrial and commercial customers might be in a better position to take the greater risks and reap the greater rewards involved in real-time pricing.

9.5. Energy-Only Markets

9.5.1. Existing energy-only markets

While various markets have implemented some form of capacity markets, there has not been a clear indication that such capacity markets have actually met their main objective to attract new investment in resources to meet increasing demand in those markets. In contrast, as mentioned earlier, several markets in the US and abroad have been operating without facing serious capacity shortages. Moran and Skinner present a more detailed discussion of the Australian approach to ensuring resource adequacy in Chapter 11.

The markets listed in this category share a common feature, namely none of these markets have established formal capacity mechanism to date (i.e., an energy-only resource adequacy mechanism). Market-based energy prices, the sale of capacity services to RTOs, and bilateral contracts for energy are the only avenue for generators in these markets to cover their operating costs, and contribute toward the recovery of their original investment costs. For example, Australia has no capacity payments but a very high offer cap.

The offer cap of $AUS 10 000 can provide a strong incentive for resources to provide electricity service and for LSEs to maintain forward supply contracts to avoid paying thousands of dollars per MWh for covering their retail service obligation in the spot market during a shortage. The high offer caps in Australia have increased bilateral contracting between buyers and sellers, which has resulted in lower average spot market prices. In contrast, Alberta has a much lower offer cap of $C1000 with no mitigation of generator offers. New Zealand has no formal offer cap.

In Texas, after the passage of legislation in 1999 that deregulated retail competition in 2002 and as ERCOT prepared to operate as a single control area[37], the Public Utility Commission of Texas (PUCT) approved the wholesale market rules with a $1000 offer cap. ERCOT saw a rush on investment in base-load combined cycle capacity without the support of a capacity mechanism. While the reserve margin in ERCOT was sufficient in the years immediately after retail deregulation, concerns were raised that the wholesale market neither sent the appropriate market signals that valued the location (and deliverability) of that investment nor provided sufficient market signals to value the operational characteristics of generation and load resource (e.g., flexibility of real-time dispatch, minimum loading of generation units, and start-up times). In addition, almost all of the new dispatchable generation was designed to operate as base-load with little or no new peaking generation entering the market.

In deliberations on the appropriate resource adequacy mechanism to choose for ERCOT, competitive retailers and industrial consumers strongly objected to the use of a capacity market approach. A retailer group suggested even higher offer caps than the PUCT accepted, reflecting their strong dislike of a capacity payment approach to resource adequacy. At least one of the PUCT commissioners expressed reluctance to institute a potential subsidy to generation, in the form of capacity payments, which once present would be difficult to remove.[38]

In 2006, the PUCT, which regulates both the wholesale and retail markets of ERCOT, adopted a combination of market power and resource adequacy rules that explicitly

[37] ERCOT began operating as a single control area on July 31, 2001 when pilot retail competition began in ERCOT. Prior to that date, ERCOT consisted of ten individual control areas operated by each major integrated electric utility in ERCOT.
[38] Public Utility Commission of Texas, Project No. 24255, *Rulemaking Concerning Planning Reserve Margin Requirements*, Memo from Commissioner Barry T. Smitherman, 15 July 2005.

rejected capacity payments in favor of raising the system-wide offer cap to ensure resource adequacy.[39] In its resource adequacy rule, the PUCT stated that it adapted the Australian energy-only resource adequacy mechanism to the ERCOT market.[40] In this proceeding, the Commission adopted Substantive Rules 25.504 (Market Power) and 25.505 (Resource Adequacy). The combination made ERCOT an energy-only market, in contrast to a capacity-and-energy approach used in electricity markets in the Eastern Interconnect of the US.

In the ERCOT resource adequacy mechanism, the offer cap is to be raised from the $1000 level that prevailed when the rule was adopted in August 2006 to $3000 in 2009. In addition, the Commission ended a system-wide market mitigation measure, the Modified Competitive Solution Method (MCSM)[41], which changed market clearing prices *ex post* under certain market conditions that suggested economic or physical withholding might have occurred. The PUCT also expressed its intention to rely on increased market-based demand response to meet its resource adequacy goals.[42] Increased market-based demand response also would weaken the potential for market power abuse during times when scarcity pricing was expected.

As part of this rulemaking project, the Commission developed a formal definition of market power, reduced mitigation on smaller market participants, and gave larger market participants the opportunity to apply for the Commission's approval of voluntary mitigation plans. The rule will raise the offer cap in ERCOT-procured markets to allow generation and load resources the opportunity to recover their fixed costs, improve incentives for bilateral contracting, and increase the transparency of ERCOT-procured ancillary service and energy markets. This is a courageous step by the Texas Commission that is regarded as politically infeasible not only in California, which is still recovering from the 2001 energy crisis, but even in MISO that has recently opted for an energy-only approach.

9.5.2. Market power abuse and scarcity pricing

Because price spikes could be substantially higher under an energy-only resource adequacy mechanism with medium-to-high offer caps, regulators would have an even greater need to distinguish the difference between scarcity pricing and market power abuse. Therefore, there is a need to supplement this *ex ante* framework with substantial market monitoring resources and greater transparency to proactively address potential market power abuses.

Various markets that are using the energy-only approach to resource adequacy have used the following approaches to address potential market power abuses.

9.5.2.1. Transparency of offers into ISO-procured energy and ancillary services markets

Increased disclosure will increase market transparency, providing incentives for market participants to make offers into ISO-procured energy and ancillary services markets that are consistent with the properly functioning competitive market and not the result of market power abuse or other market manipulation. The implementation schedule for

---

[39] Public Utility Commission of Texas, Project No. 31972, *Order Adopting Amendment to Substantive Rule 25.502, New Substantive Rule 25.504, and New Substantive Rule 25.505*, p. 6. A more detailed discussion of the development of this rule can be found in Schubert et al. (2006).
[40] *Ibid*, p. 42.
[41] See Hurlbut et al. (2004).
[42] Public Utility Commission of Texas, Project No. 31972, *Order Adopting Amendment to Substantive Rule 25.502, New Substantive Rule 25.504, and New Substantive Rule 25.505*, pp. 68–9.

disclosure is also being tied to the schedule for increases to the offer cap, thereby further emphasizing the PUCT's decision that these two issues are interrelated.[43]

The interrelationship the PUCT cites is consistent with disclosure policies in electricity markets in the United States and other foreign markets. In FERC jurisdictional markets, for instance, resource-specific information submitted into an ISO-procured market is released 6 months after the information was gathered, which is consistent with heavily mitigated individual resource offers and a low offer cap.[44] Quick disclosure of resource-specific information appears to provide limited benefit under these circumstances, because market participants are protected *ex ante* from potential price spikes, know the limited range in which the offers are made, and rely on mechanisms run by an ISO to trigger scarcity pricing in the markets.

In contrast, an energy-only resource adequacy mechanism with lighter price mitigation and high system-wide offer caps work best when the ISO discloses resource-specific offers or unit output quickly. This approach is based on the experience of the Australian market and the belief that companies that have the potential of abusing their market power will be reluctant to expose themselves to public criticism resulting from actions they take in the market to raise prices. This combination of lighter mitigation and quicker disclosure is seen in established electricity markets outside of the United States: the Australian electricity market discloses resource-specific offers with the names of the generators making the offers within 24 hours; the New Zealand electricity market discloses the same information within 14 days and may shorten the disclosure window in the near future; and the Alberta electricity market displays the output of each generator, by name, on its website in real-time.

9.5.2.2. Structural components to address market power

The Texas Legislature has put in statute a limitation on ownership of no more than 20 percent of installed capacity in the ERCOT market. In addition, the deregulation of the Texas market required integrated utilities that would offer customer choice to unbundle into separate retail, wires, and generation companies with strict code of conduct rules limiting their interaction.

9.5.2.3. Focus market monitoring on key players

In ERCOT, the PUCT instituted a provision in its market power rule that stated that market participants that owned or controlled less than 5 percent of total installed generation capacity would be deemed not to have system-wide market power (informally known as "small fish swim free" provision). As a result, the market monitor can focus more time and energy on the actions of a handful of large players in the market who would be considered most likely to be pivotal suppliers during non-peak system conditions and have the potential of exerting unilateral market power.

9.5.2.4. Voluntary mitigation plans

In ERCOT, market participants with portfolios larger than 5 percent of the installed capacity have the opportunity to earn scarcity rents on their units. If they are uncertain whether

---

[43] Public Utility Commission of Texas, Project No. 31972, *Order Adopting Amendment to Substantive Rule 25.502, New Substantive Rule 25.504, and New Substantive Rule 25.505*, pp. 27–8.

[44] Recently, a number of these ISOs have decided to implement scarcity-pricing mechanisms that prescribe specific situations when prices can rise to the offer cap.

the offers from those units would be considered an exercise of market power, they have the opportunity to apply for a voluntary mitigation plan with the PUCT. The PUCT would review the plan and, if approved, would provide the market participant with an absolute defense against charges of market power abuse as long as it adhered to the voluntary mitigation plan.[45]

9.5.2.5. Keeping market participants informed on short-tem overall supply and demand
Market participants are provided short-term forecasts to assist them with their unit commitment decisions (quick-start generation and demand resources). In the Australian market, for instance, the market operator emphasizes the importance of Projected Assessments of System Adequacy (PASA) in informing market participants of unit availability and load forecasts.

9.5.3. Demand participation

One of the current problems in restructured electricity markets is the highly inelastic demand for electricity among all but the largest consumers. Zarnikau covers demand elasticity and demand response in Chapter 8. Such inelastic demand requires a substantial *generation* (as opposed to *resource*) reserve margin.[46] When the generation reserve margin falls, an electricity market is vulnerable not only to high prices (e.g., scarcity pricing) but also to abuse of market power, because differentiating between scarcity pricing and market power abuse is difficult to prove after the fact. Increasing market-based demand response increases competition with generation at near-peak or at peak demand, reducing the need for or scope of *ex ante* mitigation of potential system-wide market power abuse while allowing for scarcity pricing.

An energy market-based resource adequacy mechanism would require widespread and active participation of demand-side resources, which could be encouraged through access to interval metering for residential and small commercial customers through the deployment of advanced “smart” meters. In May 2007, the PUCT adopted an advanced metering rule for ERCOT for this very purpose.[47] Such a mechanism increases the rewards for demand-side participation, and as a result allows the market to find and use the services of end-use customers with a lower VoLL and a greater willingness to reduce electricity in response to high prices. As a result, the market can avoid involuntary curtailments with a lower system-wide offer cap. Encouraging “reliability at a price” through voluntary, price-sensitive load shedding (curtailable load) would allow market participants to more efficiently reflect their value of reliability while maintaining the overall reliability of the grid.

The amount of curtailable load deployed in a given year would be a function of anticipated prices and would complement the entry and exit of generation resources in a generation building/retirement cycle. Demand-side resources that are available for deployment when prices are high because of a low generation reserve margin could serve as a shock absorber for end-use customers in the face of the time lag of building new power plants.

---

[45] The Texas Legislature considered legislation that would make this voluntary mitigation provision mandatory for the two largest generation portfolios in ERCOT, however, the proposal failed to receive approval before the end of the Legislative Session.
[46] Interruptible customers provided a resource in regulated markets.
[47] PUCT Substantive Rule 25.130, *Advanced Metering*.

The thinner the generation margin, the higher market prices would be, which would provide more demand response. As more generation enters the market, it increases the generation margin, lowers market prices, and reduces the benefits for participation of demand resources in the market during the years with abundant capacity.

Retailers will compete on managing price risk, with the potential of more customized packages offered to end-users. The result should be a more vibrant retail market. All parties will benefit from less generation capacity with limited use standing by. The same level of economic activity will be served by less generation capacity without sacrificing reliability of customer who places a high value on continuous service. Load management or "peak shaving" programs, which focus on reducing electricity use during summer afternoons, are more effective when prices are predictably high for enough hours to justify an investment in peak-shaving technologies or processes. Zarnikau presents a more detailed discussion of this issue in Chapter 8.

9.5.4. Limiting excessive wealth transfers

The time lag between the market price signal and the entry of new generation complicates market mitigation that, if unchecked, could lead to significant transfers of wealth to generators. Therefore, the newly restructured wholesale electricity markets should have some limit on the earnings associated with very high offer caps to ensure scarcity pricing without price gouging. For instance, the Australian approach also includes a backstop feature called the cumulative price threshold (CPT), which caps the cumulative fixed cost recovery over a 1 week period. When the limit is reached, the resource's $AUS 10 000 offer cap drops to $AUS 50–$AUS100/MWh for no more than a week. The threshold that triggers the drop in the price cap has not been reached more than once in a year.[48]

In contrast, the ERCOT wholesale market has lower offer cap, but limits the amount a resource can capture on an annual basis to $175 000 per MW. When that limit is reached, a much lower offer cap applies for the remainder of the calendar year.[49] Having such a cumulative cap in these markets with relatively high offer caps, however, could create a bright line, which enables pivotal suppliers to collect the allowed rents while staying within the allowed limits. Any market mitigation approach would need to address the problem of pivotal suppliers in the markets.

Energy-only markets can work when two types of issues are addressed: transform the electricity market into a commodity market and address potential market power abuse through a combination of structural remedies (including transparency) and market monitoring tools. Concerns about free riders may require contracting or capital requirements (margin calls).

9.6. Transitional Mechanism Based on Energy-Only Markets with Contracting Obligations

Resource adequacy mechanisms based on energy-only markets are premised on the notion that consumers are interested in buying deliverable energy, and hence generation capacity is valued based on its availability to produce energy at a given price. Under such

[48] Authors' communication with Peter Adams, Manager, Surveillance and Enforcement, Markets Branch, Australian Energy Regulator, February 1, 2005.

[49] The lower offer cap in ERCOT market is $500/MWh or 50 times Houston Ship Channel Natural Gas Price Index, whichever is larger.

Els UK Job: CDI ch09-i047172 16-11-2007 4:15p.m. Page:355 Trim:165×240MM Float:Top/Bot TS: Integra, India

mechanisms, generators can recover their investment costs either through very high spot prices that reflect scarcity rents or through long-term bilateral energy contracts with load-serving entities that specify a fixed market-based energy price at levels significantly below scarcity pricing but sufficiently high to cover generator's amortized fixed costs. These bilateral contracts are essentially energy market-based call options on generation capacity, with the capacity cost premium comparable to the premium on a financial instrument that has a set strike price (i.e., the fixed price for deliverable energy). Such mechanisms are contrasted in this chapter with resource adequacy approaches that are based on introducing an artificial capacity product for which there is no intrinsic demand. Capacity-based resource adequacy mechanisms can take the form of a direct capacity payment, as is the case in several Latin American countries, in European countries such as Italy and Spain, and in South Korea. Alternatively, the regulator can specify through regulatory fiat the needed quantity for capacity based on engineering considerations or specify an administrative demand curve for capacity (e.g., NYISO) from which the capacity price is inferred through a mitigated, auction-based procurement process. As discussed earlier, capacity payments, whether derived from an administrative demand curve or imposed directly, are aimed at restoring the missing money given the prevailing energy market and hence can complement any energy market design regardless of any pricing inefficiency.

On the other hand, recovery of investment cost through energy-only adequacy mechanisms, whether the amount of installed capacity is implied by customer preferences for reliability or dictated by engineering standards and load forecasts, requires that energy prices be efficient and be allowed to reflect scarcity and value of unserved load. In addition a functional market for risk must exist, either over the counter or centralized, which will enable generators and loads to trade investment and price risk. Such a market is essential for protecting customers and LSEs against extreme price excursions and for smoothing out boom–bust cycles that may adversely affect the cost of capital for investors. The restructuring of the electricity industry has focused primarily on improving efficiency of energy spot markets and mitigation of market power. Little attention has been devoted to facilitating markets for risk, which become crucial in an energy-only market where prices are allowed to reflect scarcity. The presumption is that fear of exposure to high spot prices will drive loads to seek cover of bilateral contracting arrangements.

However, such a leap of faith leaves some open questions:

- An LSE to contract forward for 100 percent of its forecasted peak load is not optimal given load uncertainty and potential load migration.
- Credit constraints may prevent small retail energy providers from entering into contracts that extend far into the future.
- There is always the option of bankruptcy which, like it or not, provides a hedge against extreme low-probability events. The question is then, whether investors in new generation are willing to take the risk of bankrolling an irreversible large-scale investment with a 30-year cost recovery horizon based on relatively short-term contracts and for only a portion of the capacity.[50]
- Can the system operator entrusted with maintaining reliability count on such investment or on load response in the face of a tight supply and demand balance?

[50] This concern is not limited to the electric generation industry. For instance, the development of new sources of oil and gas involve large capital costs, with the output of these investments subject to volatile and uncertain prices over the life of the investment.

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T.W:30PC open recto 1 Color 49 Lines

Currently, there are two markets in the US, MISO, and CAISO, where resource adequacy is assured through market-based bilateral contracting obligations imposed on loads or LSEs, described below.

MISO: In February 2007, the Midwest ISO (MISO) filed with FERC a resource adequacy proposal that relies on energy-only remuneration and contracting obligations imposed on LSEs. The proposal retains the $1000 per MWh energy offer caps during non-scarcity conditions (i.e., when reserves are not deployed for energy production). However, an administrative demand curve for reserves will be used to set reserve prices during scarcity conditions when operational reserves are deployed. These reserve prices will also be added to the energy clearing prices during such scarcity periods. The demand curve for reserves, which will be used in the day-ahead and real-time markets, will allow scarcity pricing to rise as high as $3500 per MWh, which is MISO's estimate of VoLL. Real-time co-optimization of energy and reserves will be implemented to improve resource utilization during scarcity conditions.

The state commissions within MISO would be expected to enforce a contracting requirement for all loads, both traditional cost-of service load and competitive retail loads, to ensure resource adequacy. The desirability of a "must-offer" availability requirement in day-ahead markets for contracted resources in MISO will be reviewed in the future.

CAISO: In California, the unpleasant experience of the energy crisis led the California Public Utility Commission (CPUC) to take a proactive role in assuring that LSEs enter into bilateral contract that will ensure adequate resources to meet local reliability requirements determined by the California Independent System Operator (CAISO). The low offer cap of $400/MWh in the CAISO energy market and the absence of any capacity payments made bilateral contracting obligations an essential backstop mechanism, which would reduce the CAISO's heavy reliance on RMR contracts to meet local reliability needs. In a series of orders, the CPUC set the ground rules for a resource adequacy requirement (RAR) program based on mandatory bilateral contracting obligations imposed on CPUC-jurisdictional LSEs.

The initial CPUC Resource Adequacy Orders were issued in 2004[51] where the CPUC established the LSE obligation framework, established qualifying capacity rules, and authorized a wide range of resource types. In a subsequent order in 2005[52], the CPUC clarified the notion of monthly capacity versus peak load, established required elements for standardized contracts, and clarified the availability obligation to the CAISO of contracted generators. In 2006 the CPUC issued another order[53] that resolved a number of regulatory uncertainties including treatment of forced outages versus scheduled outages, title clearing, creditworthiness, and the role of intermediaries.

The order also modified the required elements of tradable, standardized capacity contracts, and authorized trading via bulletin boards or exchanges. The contracting obligations imposed on the LSEs are based on load forecasts developed by the California Energy Commission (CEC) and on local reliability needs determined by the CAISO. There are nine local reliability areas and each LSE is required to carry contracts or own generation covering 115–117% of its peak load share for the upcoming year in each local reliability area[54]. The contract portfolios of the LSEs are subject to compliance verification by the

[51] CPUC Orders D.04-01-050 and D.04-10-035.
[52] CPUC Order D.05-10-035.
[53] CPUC Order D.06-07-031.
[54] To increase liquidity and facilitate the process the local reliability areas in northern California have been aggregated into two procurement areas.

CPUC[55]. In August 2006 the CPUC launched a proceeding aimed at determining whether there is a need for a centralized capacity market that will supplement or replace the current contract-based resource adequacy approach, which has several shortcomings.[56] One of the problems with the current program is the overwhelming burden of compliance verification due to the diversity of contracts.

Furthermore, the dual role of the contracts addressing both price hedging and local reliability needs is problematic since from a financial perspective it makes little sense to contract forward for over 100% of forecasted annual load. Optimal hedging of variable load would normally cover only a fraction of the peak load with forward fixed price contracts while the remainder is served through spot market procurements. Covering 100% of the peak load with forward contracts is suboptimal, since under such a strategy the LSE is over hedged when demand is low. The LSE, as a result, would have to sell its excess energy in the balancing at a potential loss, since low demand in the face of oversupply is typically correlated with low spot prices.

Call option obligations imposed on wholesale customers and LSEs are intended to overcome some of these difficulties and serve as a backup resource adequacy mechanism that corrects for possible failure in the market for risk with minimal intrusion in private risk management practices. They provide the same function as capacity products, but their pricing and performance obligation are linked to energy markets that must, therefore, be efficient and unmitigated. Ideally, spot price exposure should motivate market participants to manage their exposure to price and volume risk through a portfolio of forward contracts, call and put options, and spot trading. However, market imperfections, spot price distortions due to regulatory interventions and reliability-motivated, out-of-market operator actions, result in over-reliance on the spot market by buyers, cost recovery shortfalls by producers, and under-investment.

Such misallocation of risk in the electricity supply chain creates a vicious cycle where underhedging by load exposes customers to price risk, resulting in regulatory intervention to suppress price spikes. But such suppression of spot prices creates shortfalls in generators cost recovery and reduces incentives for investment, which in turn may result in higher spot price volatility due to shortages and threaten reliability. Standardized call option obligations can fulfill the role of a capacity mechanism in a transitional energy-only market,which has not matured to the point where voluntary contracting, financial intermediation, and load response can meet local reliability needs.

9.6.1. Call options as mandatory load hedging

Proposals for various forms of contracting obligations, in the form of physically covered call options, are described by Oren (2000, 2005a, b) and by Vázquez et al. (2002). Mandatory load hedging, although with no physical cover requirements, has also been proposed

[55] In a 2004 long-term procurement decision (D. 04-04-003) the CPUC assigned the responsibility for new generation capacity buildup to the investor owned utilities (IOUs). The rule empowers the IOU to enter into contracts with new generation capacity and allocate the capacity cost component of the contracts, for up to 10 years, to all CPUC-jurisdictional LSEs in the IOUs service territory. The rule opens the door for an auction-based mechanism for reselling the energy tolling rights to a contract so that the capacity cost component can be inferred by deducting the energy value from the total contract cost.

[56] The above summary is based on a presentation by Mike Jaske of the CEC at a CAISO Market Surveillance Committee Meeting held on August 8, 2006.

recently by Hogan (2005, 2006) in the context of an energy-only market proposal. A recent implementation of a call option approach with a central procurement auction in Brazil is described by Bezerra et al. (2006), and a similar approach has been proposed in Colombia by Cramton and Stoft (2007). As mentioned earlier, a mandatory contracting obligation has also been adopted in its recent MISO proposal to FERC and in California under a resource adequacy rule, which mandates that LSEs cover their peak load and reserve margin through bilateral contracts that are subject to verification.[57]

Contracting obligations are intended to greatly reduce the potential for a boom–bust cycle by effectively imposing mandatory price insurance as a way to protect customers against high spot prices instead of artificially suppressing these prices. For such contract to be properly priced and to address the "missing money" problem, unhedged spot prices for energy and operating reserves must be allowed to fully reflect scarcity conditions. The purpose of mandatory call options is to restore the price protection offered to customers through price caps and ensure resource availability at these prices with minimal interference in the private risk management markets. This result can be achieved by setting the strike price for the mandatory call options sufficiently high (say $1000 per MWh, like the prevailing price cap in many markets). Under such a scheme any LSE has to cover its peak load and appropriate planning reserve requirements with the standardized backstop call options or any forward contract or call options of comparable duration and equal or lower strike price. The call options give the holder the right but not the obligation to obtain a fixed amount of energy at the strike price, which can come from a generation resource or a curtailable load resource.[58] Holding options allow LSEs to buy cheaper energy from the spot market when available, reducing the cost to consumer as compared to forward contract that entails a "must-take" obligation at the contract price. The LSE benefits from holding call options, which offer customers the same protection as a price cap without locking them into a fix price.

Further price protection can be obtained through traditional private contracting. Call options can be obtained from generation resources with verifiable physical capacity to cover the contract or from interruptible firm load that by selling a call option commits to curtailment when the spot price reaches the strike price. The opportunity for load to sell call options effectively gives load an opt-out opportunity, since the revenue from selling the call option will offset their share of call option obligation cost that they may be subject to under an LSE. Resources that are not bound by call option contracts are allowed to offer their energy into the spot market at prices that exceed the strike price, while generators and interruptible loads that have offered call options are obliged to offer the corresponding energy into the spot market at the strike price or below and are liable, in case of non-performance, for the difference between the uncapped spot price and the strike price. Such liability automatically implements a non-performance penalty scheme based on market economics that keeps the customers unharmed.

---

[57] CPUC orders D.04-01-050, D.04-10-035, D.05-10-035, D.06-07-031.

[58] An ISO might find it prudent to require LSEs to procure some curtailable load resources with call options that have a strike price greater than $1000 per MWh in order to maintain system reliability during extreme weather events or random emergency conditions. Such an approach would allow the ISO to avoid mandatory load shedding by having loads that want to consume electricity during emergency conditions to pay other loads to curtail. The higher strike prices would reduce the upfront cost of procuring such voluntary load shedding. Such contracts could be procured annually rather than for 3 years to increase the potential range of load participation.

9.6.2. Centralized versus decentralized market

Backstop call option obligations can be met through bilateral contracting with generators or interruptible customers; however, in order to achieve the resource adequacy objective and provide incentives to new investment in generation capacity, it is necessary to define the call options with at least a 3 year forward lead time and an obligation duration of one to several years. The need for such a lead time has been recognized in other mechanisms mentioned earlier such as the PJM RPM, CRAM, and the ISO-NE FCM, as a way to allow new entrant with no "iron in the ground" to participate in the market and offer call options against capacity that they intend to build. Such a forward-looking approach is essential to mitigate market power of the incumbent generators in the call option market. The problem with forward-looking commitments is that load forecasts change, load migrates among LSEs, and small REPs may not have adequate credit to cover such long-term obligations.

These difficulties may be addressed by creating a centralized market of last resort for call options. Such a market that will complement the traditional over-the-counter bilateral contract market can be managed by the ISO that will act as counterparty and assume all the credit risk. The ISO will centrally procure the backstop call options on behalf of the system load through an auction mechanism. Payment to the sellers of the options will be made at performance time and the cost will be recovered from the load through their LSE on a prorated basis. Hence load migration problem and load forecast errors are dealt with based on realized load.

Self-provision in such a centralized market can be handled in two ways. One is to allow LSEs to submit proof of contracts they hold in lieu of their call option obligations. The better approach, however, is to have holders of qualifying contracts sell call options into the centralized auction covered by the contracts. The proceeds they will receive from such sold options will offset the charges for their obligation.

9.6.3. Deliverability issues

To assure that the resource adequacy objective is achieved, i.e., sold call options translate into physical capacity, the call option must have physical cover, i.e., a seller of a call option must identify physical generation capacity or firm interruptible load or make a commitment to build the capacity within a certain time frame. The question that arises is how to guarantee that the capacity will be built where it is needed so that the energy it produces is deliverable. The approach that has been adopted in the Northeast ISOs is to define locational capacity markets that reflect transmission constraints. As a result, capacity prices in such systems will vary based on location. The call option approach enables such locational differentiation naturally in a locational marginal pricing (LMP)-based system, since the financial liability associated with a call option and the opportunity cost of committing to the strike price are determined by the difference between the LMP and the strike price so that the call option premium will vary by location.

There is still the question, however, of how granular the call option obligation and the corresponding physical cover should be. The physical cover associated with the call options can have the same granularity as the financial obligation or coarser. For example, the call option obligation can be zonal, implying that non-performance penalties will be based on the difference between the average zonal spot price and the strike price, but the physical cover can be anywhere in the system. In such a case, the generator selling the call option can cover its financial risk with capacity anywhere in the system and short-term transmission congestion rights between the location of the capacity and the location corresponding to the call option the generator sold.

9.7. Conclusions

Workably competitive electricity markets with robust risk management schemes do not require regulatory-based capacity payments to generators. While this proposition was supported in the abstract by US academics and policymakers, the institutional realities of piecemeal electricity deregulation have made it seem unrealistic to many in the US. Working examples of such a market, such as those in Australia and Alberta, were either ignored or dismissed as anomalies.

For the past 10 years, the worldwide debate about key elements of wholesale market design has been most intense in the US. The US debate has been dominated by two market design elements – real-time dispatch and resource adequacy. In both cases, PJM, the ISO-NE, and the NYISO took a strong, clear position in that debate – a combination of nodal real-time dispatch with ICAP markets (capacity resource adequacy mechanisms) – which carried over into the market design debates at FERC and in three other US ISOs: CAISO, ERCOT, and MISO.

Stakeholders and the public utility commissions in CAISO, ERCOT, and MISO debated long and hard on the benefits and drawbacks of the nodal/capacity market combination adopted in the eastern US. These markets adopted the nodal real-time dispatch but rejected the ICAP approach. The rejection of capacity mechanisms was the result of the failure of existing capacity mechanisms to meet their primary objective of encouraging new investment in generation and their troubled evolution into less market-friendly, more complicated forms.

As a result, the momentum in resource adequacy has shifted toward the energy-only approach, something almost unimaginable a few years earlier. While the energy-only resource adequacy mechanism is the mechanism of choice from an economic theory perspective and has been proven feasible in various parts of the world, it may not be the right choice for everyone. Ultimately, the choice depends on existing circumstances of the market's infrastructure, stakeholders' preferences, and the extent to which the political environment can create the needed circumstances conducive to an energy-only approach. The success or failure of the energy-only resource adequacy mechanism at one or more ISOs will depend on the ability of those markets to put all the market design elements into place to transform their wholesale markets into electricity commodities markets, as has been done successfully in the Commonwealth markets such as Australia.

Acknowledgment

The opinions expressed in this article are those of the authors and do not represent the opinion of the Public Utility Commission of Texas or its Staff. The authors would like to thank Jess Totten, Director of Competition Division at the Public Utility Commission of Texas, for his invaluable comments on earlier drafts of this chapter.

References

Alberta Department of Energy (2005). *Refinement Options for Alberta's Wholesale and Retail Electric Markets*. Alberta, canada. 10 March.

Barroso, L.A., Rudnick, H., and Hammons, T. (2006). Second wave of electricity market reforms in Latin America.Presented at *IEEE PES Panel Session*, *IEEE General Meeting*, Montreal, Canada.

Bezerra, B., Augusto Barroso, L., Granville, S., et al. (2006). Energy call options auctions for generation adequacy in Brazil. *Proceedings of the IEEE Annual Meeting*, Montreal, Canada, June.

Boiteaux, M (1960). Peak load pricing. *J. of Bus.*, **33**,157–79.

Charles Rivers and Associates [Ruff, Larry] (2004). *A Transitional Non-LMP Market for California: Issues and Recommendations*.

Cramton P. and Stoft, S. (2005). A capacity market that makes sense. *Elec. J.*, **18**(7), 43–54.

Cramton, P. and Stoft, S. (2006). The convergence of market designs for adequate generating capacity. University of California Energy Institute Working Paper, 25 April.

Cramton P. and Stoft, S. (2007). Colombia firm energy market. *Proceedings of the 40th Hawaii System Science Conference (HICSS40)*, Big Island, Hawaii, January.

De Vries, L.J. (2004). Securing the public interest in electricity generation markets, the myths of the invisible hand and the copper plate. Ph.D. dissertation, Delft University of Technology, Faculty of Technology, Policy and Management,Available at: http://www.tbm.tudelft.nl/webstaf/laurensv/LJdeVries_dissertation.pdf.

Electricity Consumers Resource Council [ELCON] (2006). *Today's Organized Markets – A Step Towards Competition or an Exercise in Re-Regulation*? 4 December. Available at: http://www.elcon.org/Documents/Publications/12-4piom.pdf

Federal Energy Regulatory Commission (2002). Report of PJM Interconnection, LLC Re-Supporting Seasonal Capacity Commitment Structure. Docket No. EL01-63-001, 31 May.

Federal Energy Regulatory Commission (2005). *2004 State of the Markets Report*, June, Washington, D.C.

Hogan, W. (2005). On an "energy only" electricity market design for resource adequacy. Working Paper, Center for Business and Government, 23 September.

Hogan, W. (2006). On an "energy only" electricity market design for resource adequacy. Presentation at *Eleventh Annual POWER Research Conference on Electricity Regulation and Restructuring*, Berkeley, Ca, 24 March.

Hurlbut D., Rogas, K., and Oren, S. (2004). Protecting the market from 'Hockey Stick' pricing: How the Public Utility Commission of Texas is dealing with potential price gauging. *Elec. J.*, April, **17**, 26–33.

Joskow, P. (2005). Why capacity obligations and capacity markets. Available at: http://econ-www.mit.edu/faculty/download_pdf.php?id=1175

Litvinov E, Yang J., and Zhen, T. (2004). Building locational-based ICAP market in New England. Presentation at the *IEEE PES Summer Meeting*, Denver, Colorado, 6–10 June.

McNamara, R. (2006). Midwest ISO, resource adequacy in Midwest energy markets. Presentation at the *Organization of MISO States*, 8–9 May.

New York Department of Public Service (2003a). Prepared Testimony of Raj Addepalli, Harvey Arnett and Mark Reeder. Regarding a Proposal by the NYISO Concerning Electricity Capacity Pricing, Albany, NY, 6 March.

New York Department of Public Service (2003b). *Proposal Regarding Resource Demand Curve*. Albany, NY, 31 January.

New York Independent System Operator Inc. (2004a). *FERC Electric Service Tariff Article 5*, Albany, NY. Available at http://www.nyiso.com/services/documents/filings/pdf/services_tariff/services_tariff.pdf.

New York Independent System Operator Inc. (2004b). ICAP Demand Curve Review. ICAP Working Group, PowerPoint Presentation by Levitan and Associates Inc., Albany, NY, 27 May.

Oren, S. (2000). Capacity payments and supply adequacy in competitive electricity markets. In *Proceedings of the VII Symposium of Specialists in Electric Operations and Expansion Planning (SEPOPE VII)*, Curitiba, Brazil, 21–6 May.

Oren, S. (2005a). Ensuring generation adequacy in competitive electricity markets.In *Electricity Deregulation: Choices and Challenges* (M.J. Griffin and S.L. Puller, eds). (BSSEPP) Bush School Series in the Economics of Public Policy, June.

Oren, S. (2005b). Generation adequacy via call option obligations: safe passage to the promised land. *Elec. J.*, **18**, 28–42.

PennFuture (2006). RPM ain't R.I.P. *PennFu. Newsl.*, **8**(13), 25 October. Available at: http://www.pennfuture.org/media_e3_detail.aspx?MediaID=692&TypeID=3.

Schubert, E. (2005). *An Energy-Only Resource Adequacy Mechanism*. Public Utility Commission of Texas, Rulemaking Project No. 24255, 14 April.

Schubert, E., Hurlbut, D., Oren, S., and Adib, P. (2006). The Texas energy-only resource adequacy mechanism. *Elec. J.*, **19**, 39–49.

Shanker, R.J. (2003). Comments on standard market design, resource adequacy requirement. FERC Docket No. RM01-12-000, 10 January.

Vázquez, C., Rivier, M., and Arriaga, I.P. (2002). A market approach to long-term security of supply. *IEEE Trans. on Pow. Sys.*, **17**, 349–57.