One Day in Ten Years?
Resource Adequacy for the Smart Grid

Revised draft, November 2009

James F. Wilson[1]
Principal, Wilson Energy Economics

Abstract

Electric utilities and regional transmission organizations ("RTOs") in the U.S. aim to have enough electric generating capacity to meet anticipated peak loads with a "reserve margin" for reliability. The reserve margins are usually set to meet the widely-accepted "one day in ten years" ("1-in-10") resource adequacy criterion, under which the expected frequency of having to curtail firm load due to inadequate capacity should be no greater than once every ten years.

This paper evaluates the 1-in-10 criterion and the resource adequacy practices presently used to satisfy it, and proposes how these practices should be adapted for a changing electric power industry. The 1-in-10 criterion has always been highly conservative (perhaps an order of magnitude more stringent than the marginal benefits of incremental capacity can justify) and capacity planning has been even more conservative in practice. These practices perhaps make more sense for utility planners and regulatory authorities (who would have to answer for any curtailments that occur) than for the consumers who are directly affected if reliability is not maintained, but who also bear the cost of the additional capacity.

In the past, the large reserve margins resulting from conservative capacity planning practices were quickly absorbed by growing electricity demand. However, with federal and state energy policies increasingly promoting efficiency, advanced metering, and price-responsive demand (among other "smart grid" developments), peak loads will become more manageable and price-dependent, and steady growth in peak loads is no longer assured. Higher electricity prices and uncertainties about economic growth also contribute to doubts about future peak load growth.

Traditional resource adequacy practices should be adapted to better anticipate and encourage the development of price-responsive demand over the coming years; this paper suggests several changes. If not modified, current practices may lead to excess capacity that will pre-empt, discourage and delay the development of price-responsive demand that can reduce peak capacity needs and consumer costs. Excess capacity precludes the brief periods of low reserves and high electricity prices under which smart, price-responsive end uses realize most of their value; consumers could end up bearing the cost of the excess capacity, and also the cost of advanced meters and smart devices that will provide them, and the systems to which they connect, reduced value. Longer term, price-responsive demand should eliminate the need for reliability-based resource adequacy criteria and the associated capacity requirements and capacity markets.

---

[1] James F. Wilson is an economist and principal of Wilson Energy Economics; he is also an affiliate of LECG, LLC. The original draft of this paper was prepared for the 28th Annual Eastern Conference of Rutgers University's Center for Research in Regulated Industries, and the author is grateful for the comments and suggestions of several reviewers of the earlier draft. The views expressed are those of the author and do not necessarily reflect the views of any client.

One Day in Ten Years?
Resource Adequacy for the Smart Grid

Contents

I. Resource Adequacy on Our Present (Not So Smart) Grid

Electric utilities and regional transmission organizations ("RTOs") in the U.S. plan to have enough electric generating capacity to meet anticipated peak loads with a "reserve margin" for reliability. In practice, reserve margins have nearly always exceeded the target levels, and outages due to inadequate generating capacity have been quite rare.

Outages due to generating capacity shortages must be distinguished from those occurring due to distribution system or transmission system causes. Most service disruptions originate in distribution systems when storms, component failures, and various other causes lead to localized outages typically affecting a small group of customers. A recent survey found that in the United States, electric customers are without power an average of 1.3 times per year, for an average of 146 minutes per year, due primarily to problems in the distribution systems that serve them.[2] Outages in high-voltage transmission systems, by contrast, are very rare but can lead to blackouts across large areas, such as occurred in 2003 and affected 50 million people in eight states and the province of Ontario.[3] Note that while such widespread outages are more likely to occur when generating capacity reserves are low, they are not caused by inadequate generating capacity, because system operators will curtail firm loads to preserve the generating reserves necessary to operate the transmission system reliably and avoid catastrophic failures.

Maintaining adequate generating capacity entails building new capacity when and where needed, taking into account the lead times required (three years or more for major power plants). Demand-side resources, and transmission to allow distant generation to serve loads, are alternatives to new generation located near growing loads. Planning for adequate capacity resources entails determining the amount of capacity that will be needed to meet anticipated peak loads reliably, looking out several years. Peak load forecasts typically include some characterization of the range of uncertainty around a forecast median (or "50-50") value, plus a forecast of interruptible customer loads that can be reduced if needed to balance supply and demand.

A. The "one day in ten years" resource adequacy criterion

While a variety of institutional arrangements for the power industry exists in various regions of the U.S. today, the applicable resource adequacy criterion is generally the same everywhere: "one day in ten years" (hereafter, "1-in-10"). Reserve margins are planned such that the expected frequency of firm load curtailment due to inadequate capacity resources does not

---

[2] Joseph H. Eto and Kristina Hamachi LaCommare, *Tracking the Reliability of the U.S. Electric Power System: An Assessment of Publicly Available Information Reported to State Public Utility Commissions*, for the Ernest Orlando Lawrence Berkeley National Laboratory, October 2008, p. 15 Table 4.

[3] See, for instance, U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*, April, 2004 (https://reports.energy.gov/BlackoutFinal-Web.pdf).

exceed one event[4] in ten years.[5] This criterion is also expressed as a Loss of Load Expectation ("LOLE") of 0.1 events per year (equating to one event in ten years).

The 1-in-10 criterion can be expressed in terms of a more general reliability measure, the number of "Nines": a system that is available 99% of the time has "two Nines", a highly reliable system that is available 99.999 percent of the time has five Nines, and so forth (Table 1). Reliability is often measured in Nines when discussing systems that require very high levels of reliability (such as in telecommunications and information processing), and it is also used for electric power system resource adequacy in some countries. For example, the resource adequacy reliability target has been three Nines in Denmark, Finland, Norway and Sweden, while on an interconnected basis the target for the Nordel power system has become four Nines.[6] The Nines measure, which will be further used later in this paper, has three benefits: it measures reliability, rather than its lack (outages); it is unitless, facilitating comparisons to other types of reliability; and it is logarithmic, focusing attention on the orders of magnitude that are most significant when reliability is the issue. As Table 1 shows, the 1-in-10 criterion corresponds to 4.2 Nines, assuming a typical outage lasts five hours.

Table 1: The "Nines" Reliability Measure

Availability (fraction of time)	Outage hours/ year	Outage hours/ 10 years	Outages/ 10 years @ 5 hour duration	Nines
0.9	876	8760	1752	1
0.99	87.6	876	175.2	2
0.999	8.76	87.6	17.52	3
0.9999	0.876	8.76	1.752	4
0.99994	**0.5**	**5**	**1.0**	**4.2**
0.99999	0.0876	0.876	0.1752	5
0.999999	0.00876	0.0876	0.01752	6
0.9999999	0.000876	0.00876	0.001752	7

---

[4] While generally expressed as "one day in ten years", the criterion is typically interpreted to mean one outage event (of whatever duration), rather than 24 hours of outage, in ten years.

[5] See, for instance, standard BAL-502-RFC-02 of the ReliabilityFirst Corporation (the organization that sets standards applicable to most of PJM Interconnection), which calls for calculation of a planning reserve margin "that will result in the sum of the probabilities for loss of Load for the integrated peak hour for all days of each planning year analyzed (per R1.2) being equal to 0.1. (This is comparable to a "one day in 10 year" criterion)." The standard further specifies that the planning reserve margin shall be expressed as a percentage of the median forecast peak (expected to have a 50% probability of being too high and 50% probability of being too low) Net Internal Demand (net of interruptible demand). http://www.rfirst.org/ documents/Standards/RSVP/BAL-502-RFC-02.pdf.

[6] Vuorinen, Asko, *Planning of Power System Reserves*, Wartsila Technical Journal, February 2008.

The 1-in-10 criterion focuses exclusively on (involuntary) curtailment of firm customers, as opposed to reductions by interruptible customers. When electric demand rises and approaches the available capacity to generate and deliver it, interruptible customers are curtailed and other measures are taken, such as appeals to the public to conserve and voltage reductions. As a last resort, some firm electric loads will be curtailed to reduce demand to the level that can be served safely. When this occurs, the usual approach is to impose rotating outages: groups of customers sharing a local circuit are cut for a specified length of time (such as one hour) after which their service is restored, and, if the shortage persists, service to a different group of customers is interrupted.[7]

Note that the 1-in-10 criterion, by tolerating some (quite low) frequency of load loss, recognizes that attempting to ensure adequate capacity under all possible circumstances would be excessively costly, and the required amount of capacity would be difficult to determine. Both peak load and available supply are uncertain and variable, and it would be uneconomic to build capacity that would only be needed under extremely unlikely circumstances.

To identify the amount of capacity reserve necessary to meet the 1-in-10 criterion, many utilities and RTOs run probabilistic models that simulate the variability of peak loads and the frequency of generating plant outages.[8] To meet the 1-in-10 criterion, a system must have more than enough capacity to meet all demand levels in most years (roughly 9 of 10). The required reserve margins are driven not by the forecast median annual peak loads, but by estimates of future extreme peaks perhaps two or more standard deviations above the forecast median annual peaks.

Under the 1-in-10 criterion, the target reserve margins are typically around 15 percent of peak load, meaning that the total amount of installed capacity is planned to equal or exceed 115% of the forecast annual peak load reduced by the amount of interruptible load. On the large PJM Interconnection, L.L.C. ("PJM") system serving all or part of 13 states in the Mid Atlantic region, the 1-in-10 criterion leads to a 15.4% installed reserve margin.[9] With this reserve margin, the expected *available* capacity margin (taking into account typical rates of generating plant forced outages, which reduce total capacity about 7% on average) is about 8 percent.

B. Institutional approaches to achieving "one day in ten years"

Today, resource adequacy is achieved by a combination of regulatory and market mechanisms that varies from region to region. In areas where utilities remain vertically integrated (owning and operating distribution, transmission and generation facilities) and competition at the retail level has not been implemented, the utilities build or purchase generating capacity to serve their customers' future needs, with their plans and costs regulated at the state level. In regions where wholesale markets have been instituted under RTOs, the RTOs implement policies and mechanisms to ensure that adequate resources are provided.

---

[7] See, for instance, Southern California Edison Company's "Outage Center" web page, where customers can learn about rotating outages and receive real-time information about if and when their service might be interrupted. http://www.sce.com/PowerOutageCenter/Rotating/default.htm.

[8] An early discussion of this approach is found in Calabrese, Giuseppe, *Generating Reserve Capacity Determined by the Probability Method*, Transactions of the American Institute of Electrical Engineers, 1947 v. 66, p. 1439.

[9] PJM, 2009 PJM Reserve Requirement Study, September 2009, p. 2.

In the early days of restructuring of wholesale power markets, many experts and policy-makers envisioned allowing market incentives to determine what new capacity would be built, where and when. Indeed, from 1998 to 2004, following restructuring in some parts of the country, there was a boom in construction of new gas-fired generation leading to excess capacity in some areas. However, more recently the view has been that additional inducements are required to achieve the levels of capacity required to meet the 1-in-10 criterion. Shortcomings in wholesale markets, such as the lack of price-responsive demand and limits on price levels, are often cited as reasons additional incentives to build capacity are needed. However, the primary reason is that policies based on 1-in-10 call for larger reserve margins than might be provided in competitive wholesale power markets even with improved market designs and a more active demand side.

To achieve the desired reserve margins, RTOs have established mechanisms to offer inducements to retain existing and attract new capacity. The typical approach uses an auction mechanism to procure commitments to provide capacity in future periods in exchange for payments.[10] These capacity mechanisms have been highly administrative, extremely complex, and chronically controversial, and the intention has always been to phase them out as wholesale energy and ancillary services markets, and a more active demand side, further develop and begin to provide adequate reliability.[11]

The large reserve margins that result from current resource adequacy policies based on 1-in-10 result in excess capacity that tends to hold down the prices and revenues from the energy and ancillary services generating plants sell into the wholesale markets. As a result, cost recovery is increasingly through the capacity mechanisms and payments, which can become the primary source of revenue for the "peaking plants" that exist primarily to provide capacity in peak periods. For example, the parameters of PJM's Reliability Pricing Model ("RPM") capacity mechanism reflect the assumption that the reference combustion turbine peaking plant will earn only 16% of its required revenue from the wholesale markets and 84% from the capacity payment mechanism.[12]

The RTOs' capacity auction mechanisms determine payments to all capacity resources, new and existing, that are not owned by load-serving entities and have not chosen to contract in advance. The mechanisms attempt to acquire commitments from nearly all existing resources plus some

---

[10] For instance, the New England regional transmission organization, ISO-New England, operates auctions to procure commitments to provide generating capacity three years in the future, and existing and potential new sources of capacity both participate in the auctions. The details of ISO-New England's "Forward Capacity Market" (FCM) can be found at http://www.iso-ne.com/ support/training/courses/fcm/index.html. On the PJM system, the mechanism is called the "Reliability Pricing Model" or RPM. http://www.pjm.com/markets-and-operations/rpm.aspx. The New York Independent System Operator's capacity mechanism is described at http://www.nyiso.com/public/ products/icap/index.jsp.

[11] For instance, in its original order approving PJM's RPM mechanism, the Federal Energy Regulatory Commission noted that state commissions and PJM were in agreement that "capacity markets should diminish in importance to the extent energy markets in the future prove capable, standing alone, of offering adequate assurance of reliability." *PJM Interconnection, L.L.C.*, 115 FERC ¶ 61,079 (April 20, 2006) at P. 170-71.

[12] See PJM, *RPM Planning Parameters for the 2012/2013 Base Residual Auction*, identifying the "Cost of New Entry" as $112,868/MW-year and expected net revenues from energy and ancillary services of $18,585/MW-year (16%). http://www.pjm.com/markets-and-operations/rpm/~/media/markets-ops/rpm/rpm-auction-info/2012-2013-rpm-planning-parameters.ashx.

new resources; as a result, incentives to physically or economically withhold to raise capacity prices can be strong and various forms of mitigation are often imposed. Small increases in the target capacity requirement (the "demand" to be satisfied in such auctions), or small decreases in the amount of capacity offered, can significantly increase the auction clearing price paid to all resources and the total cost to consumers of the acquired commitments. Consequently, the many administrative details of these mechanisms have substantial cost impacts and are frequently a source of controversy and negotiation among stakeholders.

C. In practice, resource adequacy has been more conservative than 1-in-10

Capacity planning approaches result in resource adequacy that usually exceeds the 1-in-10 criterion. While 1-in-10 has been accepted in principle, in practice planners and regulators understandably have as a goal that curtailments never occur, and there may be "thumbs on the scale" as resource adequacy is implemented.

There are a number of ways resource adequacy in practice has often been more conservative than the 1-in-10 criterion requires. As noted earlier, because the criterion is probabilistic, probabilistic modeling is required to determine the reserve margin required to satisfy it. Such models rely on assumptions about future load growth and its variability; capacity resources and their outage rates and availability during peak periods; the amount of interruptible load available; the assistance that may be available from neighboring systems during peak periods; and the impact of actions that can be taken when reserves are low to avoid having to curtail firm customers, such as appeals to the public and voltage reductions. The tendency is often to adopt conservative assumptions for many of these values, to make the overall result of the analysis conservative (meaning, erring on the side of too much rather than too little capacity and reliability, identifying too large rather than too small a reserve margin). As a result, the resulting reserve margin may correspond to an LOLE less than, and perhaps much less than, 0.1. In addition, peak load forecasts may also rely upon conservative assumptions and err on the side of over- rather than under-forecasting future peak loads.

Utilities and RTOs can take a number of actions when reserves are low to avoid firm curtailments. These include actions to elicit maximum available generation, including both pricing and administrative approaches; recalling energy exports and requesting assistance from neighboring power systems; calling on interruptible and emergency demand-response customers to reduce loads; reducing voltage levels; and appealing to the public to reduce electricity use.[13] However, because reserves have fallen to low levels only rarely in recent years on many power systems in the U.S., there is relatively little experience with these measures and their potential impact on supply and demand. As one example, while PJM reserves 3,500 MW of transmission import capacity as a "capacity benefit margin" to rely upon under emergency circumstances, it has never actually called upon more than a small fraction of this capability.

Appeals to the public to conserve electricity can bring forth a substantial reaction. Based on analysis of actual instances of capacity shortages from around the world, the International Energy Agency concludes that a 3% reduction in electricity use can be accomplished within a

---

[13] See for instance, MidAmerican Energy Company's tariff, Rules and Regulations for Electric Service in Illinois, Section IX, Electric Energy and Capacity Contingency Plan, or PJM Manual 13, Emergency Operations,

few days through clever use of the media and other strategies.[14] Such actions provide a further buffer against the risk of curtailment of firm customers. However, calculations of required reserve margins often adopt conservative assumptions regarding the potential impacts of the actions that can be taken when reserves are low, if such actions are reflected at all.[15]

In addition, most types of power plants are most economical if built large, so new facilities will often be added that provide more capacity than needed in the first years of commercial operation. Construction times are uncertain, so the conservative approach is to target for completion before the plant is considered needed. These practices also contribute to reserve margins frequently above target levels.

Large RTOs such as PJM, ISO-New England, or the NYISO determine a single reserve margin applicable to the entire region each serves. Joint planning for a large interconnected region reduces the total reserve requirement compared to the combined result of separately-planned subregions, due to increased reserve sharing and diversity in the timing of peaks. However, joint planning for a large region also has resulted in application of the 1-in-10 criterion, which originally was applied by individual utilities, over a much larger geographic scope. One event in ten years in an area as large as these RTOs now serve is a more stringent criterion than one event in ten years for each of the several utilities within the same RTO territory.

That resource adequacy planning has been very conservative and resulted in a lower frequency of outages than the 1-in-10 criterion suggests is reflected in the reports of capacity and energy emergencies utilities submit to the North American Electric Reliability Corporation (“NERC”).[16] These reports identify the circumstances under which an outage or other emergency situation occurred, the cause of the emergency, the number of customers affected, and the amount of curtailment, if any. The NERC reports show that the vast majority of the incidents did not result in loss of load, or were caused by transmission or distribution system equipment failures, or were caused by extreme weather such as wind, snow, or hurricanes. Perhaps a dozen incidents occurred over the past decade in which there was a loss of load due to capacity shortages, the topic of this paper. Given that 134 entities file such reports, planning based on 1-in-10 would result in an order of magnitude more outages during this ten-year period.

D. Case study: What happened (and didn’t happen) on August 1-2, 2006

On July 31, 2006, in the midst of an extreme heat wave, PJM predicted that the next day, Tuesday August 1, would see a record-setting peak load on its system. It issued a news release

---

[14] International Energy Agency, *Saving Electricity in a Hurry – Dealing with Temporary Shortfalls in Electricity Supplies*, 2005, available at http://www.iea.org/Textbase/publications/free_new_Desc.asp?PUBS_ID=1481.

[15] ISO New England reflects the impacts of some emergency actions, including voltage reductions, in its installed reserve margin calculations, while PJM does not include an estimate of such impacts. ISO New England, *Installed Capacity Requirement, Local Sourcing Requirements, and Maximum Capacity Limit for the 2011/12 Capability Year,* December 1, 2008; PJM, *2008 PJM Reserve Requirement Study*, September, 2008.

[16] NERC establishes reliability standards for the electric power industry under authority granted by the Federal Energy Regulatory Commission. NERC’s *Standard EOP-002-2: Capacity and Energy Emergencies* requires entities experiencing such emergencies to file a report on each incident. The reports can be viewed at http://www.nerc.com/page.php?cid=5l65.

requesting customers to conserve electricity, especially from 3 PM to 7 PM.[17] Utilities within the PJM region also made such requests of their customers. As predicted, Tuesday was a record-setting day; PJM's metered peak was 144,041 MW on August 1 and a bit higher, 144,644 MW, on August 2 (the two highest peaks shown on Figure 1, later in this paper). Neighboring systems also experienced extreme peak loads on these days.

Never before or since has PJM's peak load reached even 140,000 MW. This peak load event was, for PJM, the "one day in ten years" discussed in this paper (it has also been referred to as a 1-in-30 year weather event). However, PJM had plenty of capacity, so it served this load with only about 1,300 MW of load reductions by its interruptible customers.

The requests to conserve electricity were not prominently communicated by the media. In the *Philadelphia Enquirer*, the article on the heat wave was in the local news section, and utility requests to conserve were briefly mentioned well into the article. The *Washington Post* ran a front page article on the heat wave, but the request to conserve was briefly mentioned several paragraphs into the article on a continuation page. Therefore, we can be sure that in many homes and businesses across the PJM region, air conditioners cooled empty rooms and lights were left on; washing machines and dryers were running whose operation could have been delayed a few hours; refrigerators were running that could have coasted a few hours; there was the usual "vampire" electricity consumption by millions of electronic devices not in use; and many other forms of electricity consumption were occurring that could have been reduced or delayed with little cost or inconvenience. Such reductions were not necessary; PJM met the extreme peak that occurred on this date without needing or receiving much help from its firm or interruptible customers, or neighboring power grids. While PJM had made a filing one year earlier with the Federal Energy Regulatory Commission stating that its resource adequacy approach and capacity market rules "no longer provide adequate assurance of continued regional reliability,"[18] it had more than adequate capacity in the summer of 2006.

II. Evaluation of the "One Day In Ten Years" Criterion

While the 1-in-10 criterion has been very conservatively applied, this section will show that it is also a very conservative criterion.

Economists have questioned the 1-in-10 criterion for many decades.[19] This section evaluates the 1-in-10 criterion from a few different perspectives, all of which suggest the criterion is at least an order of magnitude more stringent than appropriate.

---

[17] PJM Interconnection press release, *PJM Asks Consumers to Conserve Electricity on Tuesday*, July 31, 2006, available at http://www.pjm.com/Media/about-pjm/newsroom/2006-releases/20060731-h2-for-tues.pdf.

[18] PJM Interconnection, L.L.C., *Reliability Pricing Model Filing*, FERC Docket Nos. ER05-1410 and EL05-148, August 31, 2005, p. 5. The filing requested FERC to find that PJM's current capacity construct was unjust and unreasonable, and proposed the Reliability Pricing Model (RPM) as the replacement.

[19] See, for instance, Telson, Michael E. *The economics of alternative levels of reliability for electric power generation systems*, Bell Journal of Economics Vol. 6 No. 2 (Autumn 1975) p. 679; Cramton, Peter and Steven Stoft, *The Convergence of Market Designs for Adequate Generating Capacity*, April 25, 2006, p. 32; Joskow, Paul L., *Competitive Electricity Markets and Investment in New Generating Capacity*, June 12, 2006, p. 48-49; Hogan, William W., *Regulation and Electricity Markets: Smart Pricing for Smart Grids*, presentation to the Energy Bar Association Electricity Committee Meeting, October 16, 2009, pp. 21-23.

A. Economic evaluation of the 1-in-10 criterion: marginal cost and benefit

The 1-in-10 resource adequacy criterion is economically efficient if it calls for an amount of capacity that reasonably balances the incremental costs and benefits of additional capacity. Under this principle, more capacity should be built as long as its incremental cost is exceeded by the anticipated incremental benefit.

The cost of incremental capacity is the annualized cost to build and maintain the most economical type of capacity, less the amount of those costs the plant can be expected to offset through sales of energy and ancillary services in the wholesale markets. It is generally considered that gas-fired combustion turbines represent the cheapest type of capacity, and the type that would be built to meet an incremental need for capacity for reliability.

The incremental benefit of holding more capacity for reliability results from reducing curtailment due to shortages. This potential benefit depends upon the anticipated frequency of such outages (the LOLE) and the cost of outages to the electricity consumers who are curtailed (often called the "value of lost load" or "VOLL"). Additional capacity can also contribute to lower market prices for energy and ancillary services, potentially an added benefit from the consumer's perspective. However, the last increments of capacity built to satisfy the 1-in-10 criterion (or any criterion leading to a low frequency of outages) will run very infrequently and have little if any impact on these prices.

This section shows that, comparing incremental costs and benefits in this manner, the 1-in-10 criterion appears to be extremely conservative, calling for a much higher level of capacity than is justified by the economics. With an LOLE of only 0.1 outages per year (as implied by the 1-in-10 criterion), the incremental cost of capacity exceeds the incremental benefits by a wide margin across a range of reasonable assumptions. Estimates of the cost of capacity and value of service to customers suggest that a balancing of marginal cost and marginal benefit would require an outage frequency substantially greater than 1-in-10, as described in the following paragraphs.

1. The benefit of incremental capacity: the Value of Lost Load ("VOLL")

The impacts of outages on business and residential customers include loss of productivity, potential damage to electrical devices, inconvenience or discomfort due to loss of lighting, cooling or heating, spoilage of refrigerated goods, waste due to interruption of a manufacturing process, traffic snarls or accidents due to inoperable traffic lights, or missing a favorite TV show, to name just a few. Outages can also be contributing factors to injury or death. The average impact of outages on electricity customers is often quantified as the "value of lost load" or VOLL, expressed in dollars per megawatt hour of curtailed load.

Of course, the cost of an outage will be different for every customer and depend upon the customer's uses for electricity, the circumstances under which the outage occurred, the duration of the outage, whether there was any advance warning, and other factors. For some customers, a very high level of reliability is desired, reflecting the nature of the facility or uses of electricity. Many customers desiring higher electric service reliability self-provide it, by installing on-site backup generation, uninterruptible power supply (UPS) systems, or other approaches. To the extent such customers are protected from the impacts of outages in utility electric service, their

VOLLs for utility system reliability will be much lower, as suggested by a recent review of outage costs by Lawrence Berkeley National Laboratories ("LBL Outage Cost Review").[20]

In addition, utilities identify "essential use" customers who (along with other customers fortunate to share the same circuits) are exempted from rotating outages. Essential use customers typically include hospitals and nursing homes, prisons, police and fire stations, radio and TV stations, some water and sewage facilities, telephone switching stations, and emergency management and "911" systems.[21] In estimating a VOLL pertinent to rotating outages, the value of lost load for essential use customers is not relevant.

It can also be assumed that customers adapt to some extent to the level of reliability they are accustomed to receiving, and these adaptations reduce the exposure to the impacts of outages. In addition to adaptations such as self-provision of reliability, customers will be more likely to have battery backup systems for their computers, for example, or to at least set their software applications to auto-save documents, if they suffer frequent service disruptions.

A survey of the literature for ISO New England concluded that a wide range of values for VOLL, from $2,400/MWh to $20,000/MWh, can be "deemed justified by some source in the literature."[22] The LBL Outage Cost Review also suggested a very wide range, with lower values for residential customers and the highest values for small commercial and industrial customers. Other surveys have suggested similar ranges. However, outage cost surveys generally do not consider customers' exposure to outages, as information on back-up power supplies is difficult to find; customers with relatively high VOLL are more likely to have made such investments.

The Midwest Independent Transmission System Operator, the RTO for a large region of the Midwest, uses VOLL as a parameter in its ancillary services, and has recently set it to $3,500/MWh. Estimates of VOLL for the New York Independent System Operator in 2004 identified a "lower range" of $1,000 to $2,500/MWh and a higher range of $3,000 to $5,000/MWh.[23] A U.S. Department of Energy report in 2006 stated that a VOLL representing an average value in the range of $2,000 to $5,000/MWh is the "accepted industry practice."[24] Based on these and other reviews, a range of values from $2,000 to $20,000/MWh is evaluated in this paper, however, values in the $3,000 to $5,000/MWh range are considered most appropriate for this analysis.

---

[20] Ernest Orlando Lawrence Berkeley Laboratory, *A Framework and Review of Customer Outage Costs: Integration and Analysis of Electric Utility Outage Cost Surveys*, November, 2003, p. 14 (showing outage costs for large commercial and industrial customers 3 or 4 times lower if they have back-up systems).

[21] See, for instance, *Duke Energy Indiana – Essential Use Customers*, http://www.duke-energy.com/indiana-business/products/essential-use-customers.asp.

[22] Cramton, Peter and Jeffrey Lien, *Value of Lost Load*, February 2000, p. 4, available at http://www.iso-ne.com/committees/comm_wkgrps/inactive/rsvsrmoc_wkgrp/Literature_Survey_Value_of_Lost_Load.rtf.

[23] Breidenbaugh, Aaron, New York Independent System Operator, *The Market Value of Demand Response*, presented at the PLMA Fall 2004 Conference, September 30, 2004.

[24] U.S. Department of Energy, *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them*, February 2006, p. 83.

2. The cost of capacity reserve: Net Cost of New Capacity, or "Net CONE"

The marginal cost of additional reserve capacity can be represented by the annualized cost of building and maintaining the least expensive form of reliable peaking capacity, usually considered to be a gas-fired combustion turbine. Values recently developed by PJM for use in its RPM capacity mechanism can be used. While marginal capacity resources cannot expect much in the way of energy and ancillary services earnings, estimates of such net earnings reduce the cost of capacity and are also considered. The values used by PJM over the past year have ranged from roughly \$70,000/MW-year to \$130,000/MW-year on a levelized basis, with the range primarily reflecting the timing of the cost estimate (lower estimates are from 2005-2006, higher estimates from 2008) and also location.[25] Energy and ancillary services earnings are estimated to range from approximately \$10,000/MW-year to \$50,000/MW-year in various locations in recent years. This suggests a net capacity cost ("Net CONE") range from about \$40,000/MW-year to \$120,000/MW-year. While the upper end of this range reflects the more recent data (but before the impacts of the 2009 recession on these costs), PJM's capacity auction in 2008 cleared at a price close to the low end of this range.

3. Comparing the benefit of incremental capacity to its cost

If the 1-in-10 criterion is being met (the LOLE is 0.1 events/year), the last MW of reserve capacity has a 10% chance of being needed to avoid or reduce an outage in any year. To determine the potential lost load precluded by an incremental MW, an assumption regarding the average duration of an outage is required; five hours will be used. The estimated five hours duration for a typical rotating outage is based on review of hourly load shapes in a few areas of the country. Load levels tend to rise during the morning and afternoon and fall during the evening on the hot summer weekdays that are most likely to experience extreme loads in most areas of the country, suggesting an average duration of roughly five hours for a rotating outage due to capacity shortage. With this assumption, an incremental MW saves, in \$/MW/year,

0.1 (expected events/year) x 5 (hours/event) x VOLL (\$/MWh curtailed).

In general, the benefit of the last MW of capacity will equal the LOLE times the expected hours of operation during the typical outage times the VOLL. If an optimal level of resource adequacy is being provided, marginal cost equals marginal benefit, so:

Net CONE (\$/MW-year) = LOLE x hours/event x VOLL,

and the optimal LOLE = Net CONE / VOLL / 5, assuming five hours per event. Table 2 shows optimal LOLE values under various VOLL and Net CONE assumptions.

Most of the combinations of assumptions suggest an optimal LOLE in excess of one event per year. Only under the lowest Net CONE assumption and the extreme value for VOLL (\$20,000/MWh) is the implied LOLE value 0.4/year, which is still four times more frequent than "one day in ten years." Assuming a typical outage duration less than five hours would also raise

---

[25] See the RPM Planning Period Parameters for the 2011/12 and 2012/13 Base Residual Auctions, available at http://www.pjm.com/markets-and-operations/rpm/rpm-auction-user-info.aspx. The values are based on cost studies that are also available on the PJM website.

the optimal LOLE. In terms of the Nines reliability measure, the range of estimated optimal values is 2.2 to 3.6, compared to 4.2 Nines for the 1-in-10 criterion. This analysis suggests that 1-in-10 is roughly an order of magnitude more stringent than the criterion that would provide the optimal level of resource adequacy. Put another way, it would be more economical for electricity consumers for capacity to be planned such that there would be approximately one outage per system due to resource shortage per year, rather than one per decade, taking into account the impact of the outages and the cost of capacity.

Table 2: Optimal LOLEs for Various VOLL and Capital Cost Assumptions

Value of service ("VOLL")	Net Capital Cost ("Net CONE")	Hours per outage event	Optimal LOLE	Optimal Nines
$/MW-year	$/MWH	hours/event	events/yr	
$4,000	$120,000	5	6.0	2.5
$4,000	$80,000	5	4.0	2.6
$4,000	$40,000	5	2.0	2.9
$2,000	$120,000	5	12.0	2.2
$2,000	$80,000	5	8.0	2.3
$2,000	$40,000	5	4.0	2.6
$20,000	$120,000	5	1.2	3.2
$20,000	$80,000	5	0.8	3.3
$20,000	$40,000	5	0.4	3.6

B. "One day in ten years" from the customer's perspective

While "(not more than) one day in ten years" could be interpreted as a reliability pledge to each and every customer, the 1-in-10 criterion is generally interpreted as pertaining to the frequency of curtailment or load loss on an electrical *system*. However, when outages due to insufficient resources occur, typically only a small fraction of load must be curtailed to bring the system into balance. Consequently, only a small subset of customers is affected each time an outage occurs, and the frequency with which any individual customer is curtailed will be much lower than the system-wide outage frequency.

The frequency of curtailment for the average customer can be roughly estimated. If the typical outage lasts five hours, during this time an average of two percent of a system's firm load is curtailed in each hour, and the curtailment is rotated hourly, in total 10% of the customers or customer load is curtailed during each outage. These estimates are conservative and roughly based on examination of hourly load shapes; more likely, one rotating outage event would affect less than 10% of a system's customers. If it is further assumed that 50% of the customers are, or share circuits with, "essential use"[26] customers, and are therefore exempt from curtailment, the curtailment must be imposed on the remaining 50% of customers. With these assumptions, the exposed customers would be curtailed once every five outage events on average. Thus, 1-in-10 for a system translates into roughly one hour of outage every fifty years for the average customer exposed to such outages.

C. "One day in ten years" compared to distribution system reliability

While the 1-in-10 criterion might result in a risk of curtailment to the average customer of once every several decades, most electricity customers experience a much higher frequency of outages due to disturbances in the electric distribution systems that serve them – roughly two orders of magnitude (100x) higher. The comparison of the 1-in-10 criterion to distribution system outage rates also suggests that the 1-in-10 standard is extremely conservative.

Utilities summarize the number of minutes of interruption their average customer experiences with the System Average Interruption Duration Index, or SAIDI, usually expressed in minutes of outage per year. Two values are usually provided, one including all events, and a somewhat lower value excluding "major events"; the latter measures the more localized events that originate in utility distribution systems. A recent LBL report summarized utility-reported SAIDI values by census division, with major events excluded, showing a range from 107 to 212 minutes/year and a national average of 146 minutes/year.[27] The SAIDI values will of course vary for each portion of each electric distribution company's service area.

The 1-in-10 resource adequacy criterion can be expressed in minutes per year for comparison to SAIDI values, based on the rough estimates regarding curtailment quantity and duration used above. Those assumptions suggested that under 1-in-10 the average customer would be curtailed for one hour every fifty years, or 1.2 minutes per year on average. Thus, distribution system outages appear to impose roughly two orders of magnitude more minutes of outage on customers than does resource adequacy under the 1-in-10 criterion (146 compared to 1.2 minutes per year).

Electricity customers appreciate advance notice of an outage, and estimates of the duration of an outage, as this information allows them to minimize the impacts to some extent. Such information is more likely to be available under a rotating outage (a controlled event) than under a distribution system disturbance, which typically occur unexpectedly and last for an unpredictable duration. Thus, in addition to the frequency of distribution system disturbances being much greater, the impact on consumers per unit of time (VOLL) may also be greater than for the rotating outages that occur due to capacity shortages.

The comparisons of distribution system reliability and customer-level resource adequacy to the 1-in-10 criterion are summarized in Table 3. Distribution system reliability, excluding major

---

[26] One survey found that approximately 50% of customer load in California is exempt from rotating outages due to protection of "essential use" customers. Lawrence Berkeley National Laboratory, *Rates and technologies for mass-market demand response*, Paper LBNL 50626, 2002, p. 4.

[27] Joseph H. Eto and Kristina Hamachi LaCommare, *Tracking the Reliability of the U.S. Electric Power System: An Assessment of Publicly Available Information Reported to State Public Utility Commissions*, for the Ernest Orlando Lawrence Berkeley National Laboratory, October 2008, p. 15 Table 4.

events, has averaged about 3.6 Nines. By comparison, the 1-in-10 criterion corresponds to 4.2 Nines at the system level, or over 5 Nines for the customer, under the above assumptions.

Table 3: Reliability Comparisons

Types of Events (Customer or System Level)	SAIDI (minutes/year)	Availability (fraction of time)	Nines
Distribution system reliability including/excluding major events:			
SAIDI incl. major events (customer)	292	.99944	3.3
SAIDI excl. major events (customer)	146	.99972	3.6
Resource adequacy reliability:			
1-in-10 (system level)	30	.999943	4.2
1-in-10 (customer) *	1.2	.999998	5.6

* Assumes five hour duration, curtailment of 2% of peak load, 1 hour outage blocks, 50% of customers are excluded based on essential use.

D. Is planning for 1-in-10 justified to avoid the risk of more frequent outages?

Highly conservative capacity planning could reflect concern that if load forecasts or new capacity projections prove inaccurate, a region might suffer frequent, costly outages. That is, the main concern might be not one or a few outages in ten years, but the risk of unforeseen circumstances leading to multiple outages in a single year, for instance, on many hot days in a single summer. Roughly speaking, “many hot days” could be quantified as ten days in one year, or two orders of magnitude greater frequency than 1-in-10.

However, the extreme peak loads targeted by the 1-in-10 criterion occur rarely, and peak load levels on nearly all other days are considerably lower, as will be further shown in this section. Consequently, for there to be shortages on many hot days rather than just the rare extreme peak day, capacity would have to fall far short of the amount targeted under the 1-in-10 criterion. This suggests that concern about the possibility of circumstances that could lead to outages on many days does not help to justify the 1-in-10 criterion.

Figures 1 to 3 show the highest daily peak load levels attained in recent years on the PJM, ISO-New England, and California ISO systems, respectively. For PJM, compared to the two highest peak loads (which occurred on consecutive days in 2006), the next highest was 96.6% of the highest peak, and peak loads in excess of 92.4% of the highest peak (11,000 MW lower) occurred only four other times during 2005-2008.[28] Thus, even if a shortage of capacity had led to rotating outages on the highest peak days (which did not occur), the available capacity would have to have been much lower for outages to have occurred on more than a few additional days.

---

[28] While peak load data for 2009 is available, it is not shown in these figures as the peaks in 2009 were much lower due to the recession.

Figure 1: PJM Highest Daily Peak Loads 2005-2008

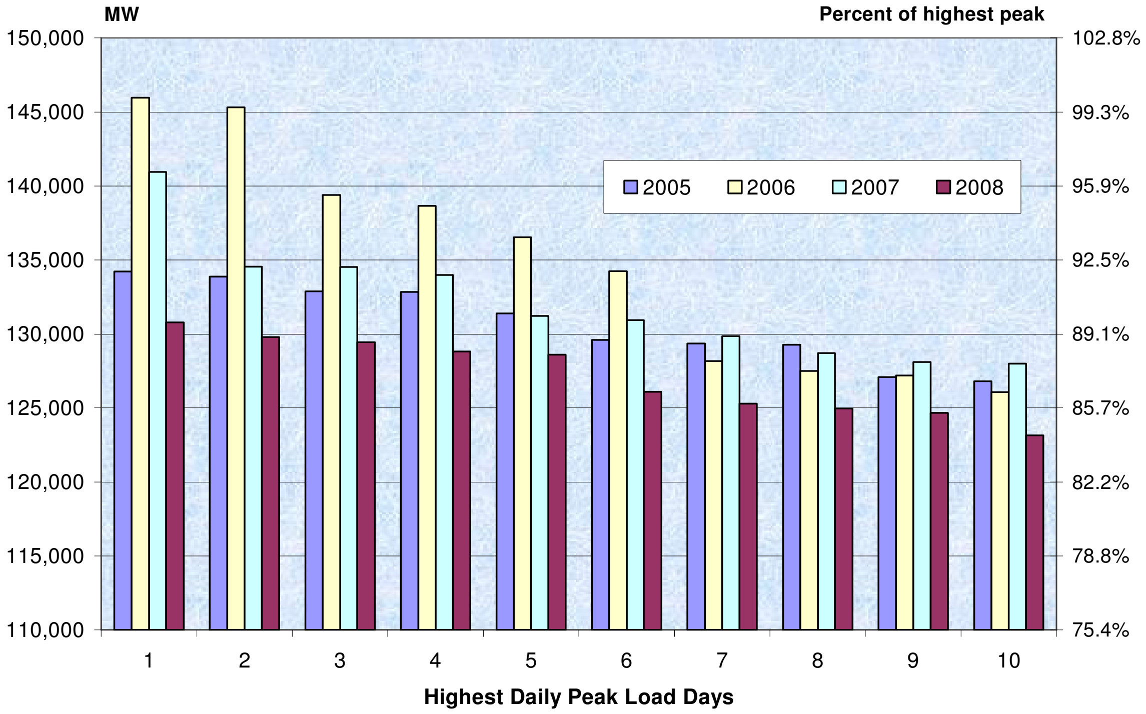
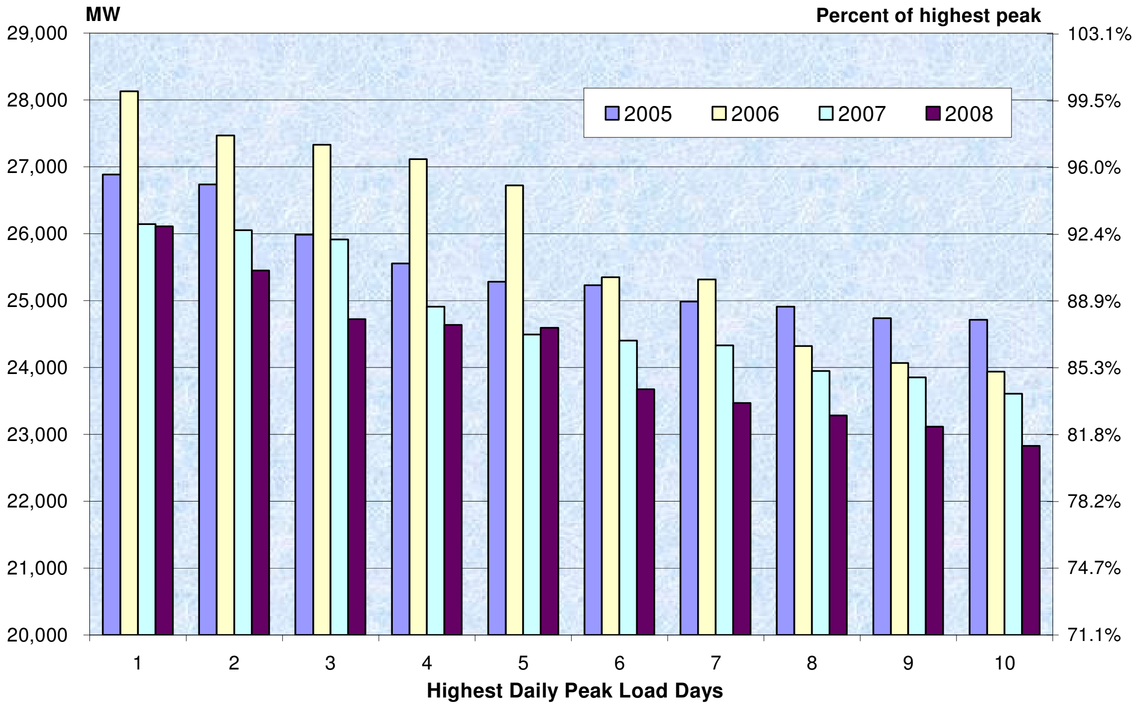


Figure 2: ISO New England Highest Daily Peak Loads 2005-2008



Similarly, on the ISO-New England system, only six additional days over four years had peak load levels in excess of 93% of the highest peak. On the California ISO system, only seven additional days over the past three years had peak load levels in excess of 93% of the highest peak. On each of these systems, the tenth highest peak in any year was typically ten percent or more below the highest peak for the year, with the exception of the mildest years when the highest peak was relatively low. To put these differences in perspective, seven percent of peak load is roughly four years of past peak load growth for ISO-NE or PJM, or five or six years of forecast growth in peak load based on 2009 projections.[29]

Models used to determine the required reserve margins to satisfy the 1-in-10 criterion provide further support for the conclusion that only a level of capacity far short of the 1-in-10 level would risk frequent outages. Figure 4 shows the relationship between the installed reserve margin and the LOLE for the PJM system, according to data provided in PJM's most recent reserve margin analysis[30] and a probabilistic model developed by the author that approximates the assumptions, structure and results of PJM's analysis.

As shown in Figure 4, this model estimates that if the installed reserve margin is approximately eight percent, far below the target of 15.4 percent, the outage frequency is one per year. Only with an installed reserve margin of under zero (that is, total installed capacity is roughly equal to the forecast median annual net peak load) would the LOLE rise to approximately 10 events per year. (That it takes such an extremely low reserve margin to anticipate 10 outages per year reflects the fact that daily peak loads close to the median annual peak level are rare, and, in addition, there is some help available from neighboring systems not reflected in the reserve margin.) Figure 4 suggests that PJM's model, which exhibits similar sensitivity to the reserve margin, would likely calculate a similar LOLE corresponding to lower reserve margins.

This analysis suggests that to risk frequent outages, installed capacity would have to be far below the level that satisfies the 1-in-10 target; it is not necessary to aim for 1-in-10 to ensure that the risk of outages on many hot days due to capacity shortages is very small. However, on much smaller systems the relationship would be different, and the anticipated LOLE would rise faster with lower levels of capacity.

---

[29] ISO New England, *2009-2018 Forecast Report of Capacity, Energy, Loads and Transmission*, April 15, 2009, available at http://www.iso-ne.com/trans/celt/report/index.html; PJM, 2009 PJM Load Forecast Report, available at http://www.pjm.com/planning/resource-adequacy-planning/~/media/documents/reports/2009-pjm-load-report.ashx.

[30] PJM, *2009 PJM Reserve Requirement Study*, September 2009.

Figure 3: CA-ISO Highest Daily Peak Loads 2005-2008

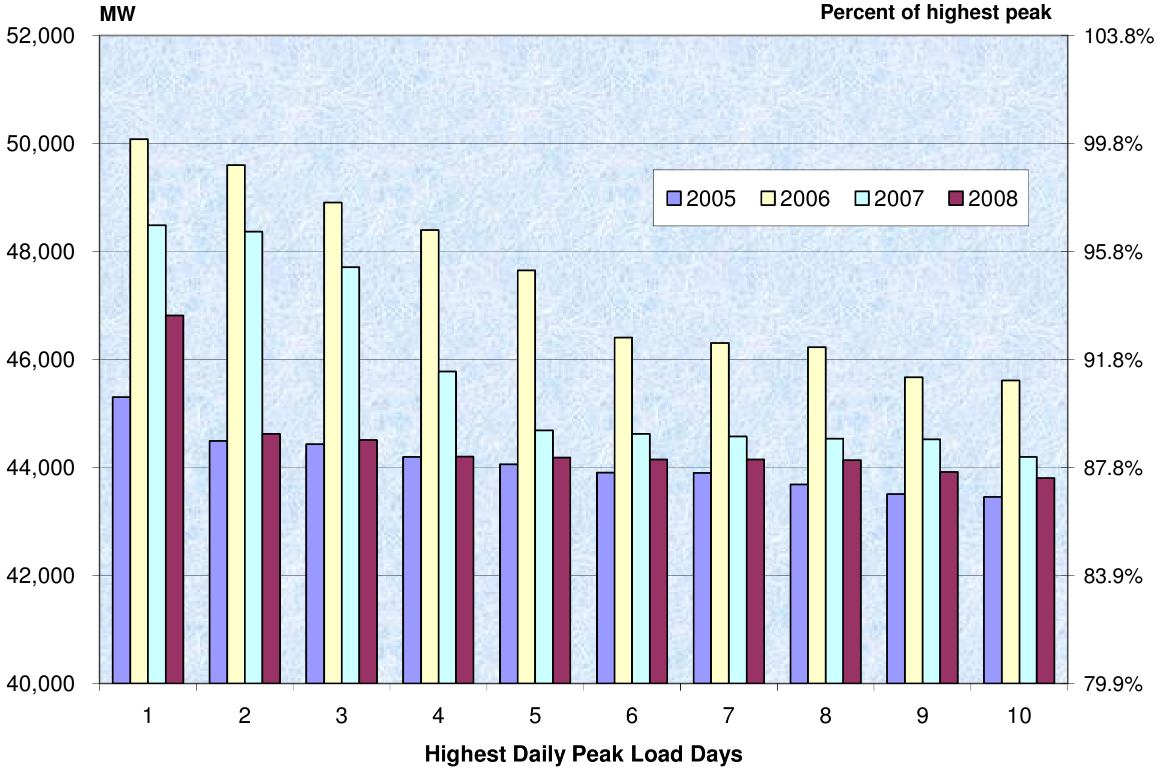
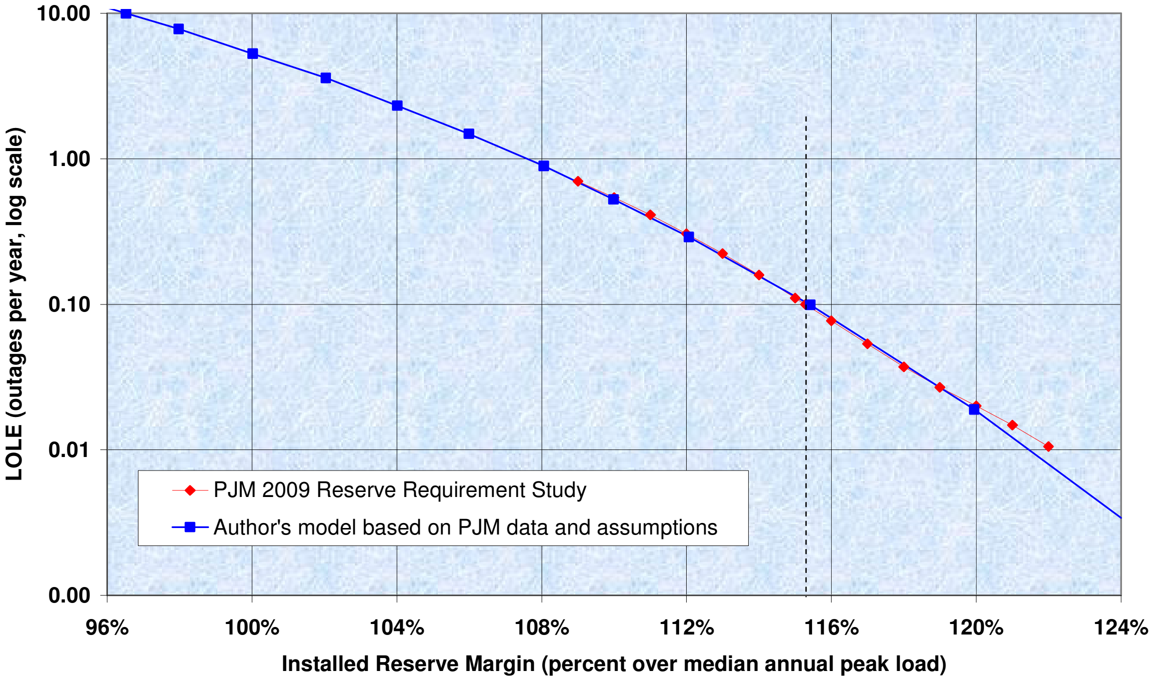


Figure 4: Relationship Between Installed Reserve Margin and Loss of Load of Load Expectation (PJM System, 2013 Data)



III. Planning for Adequate Capacity Resources: The Changing Circumstances

The preceding sections described that resource adequacy planning in the U.S. has generally rested upon a very conservative criterion, conservatively applied. Why has resource planning been so conservative? And if there was a rationale for this practice in the past, is the rationale still applicable today, and for the future?

A. Why has resource adequacy planning been so conservative?

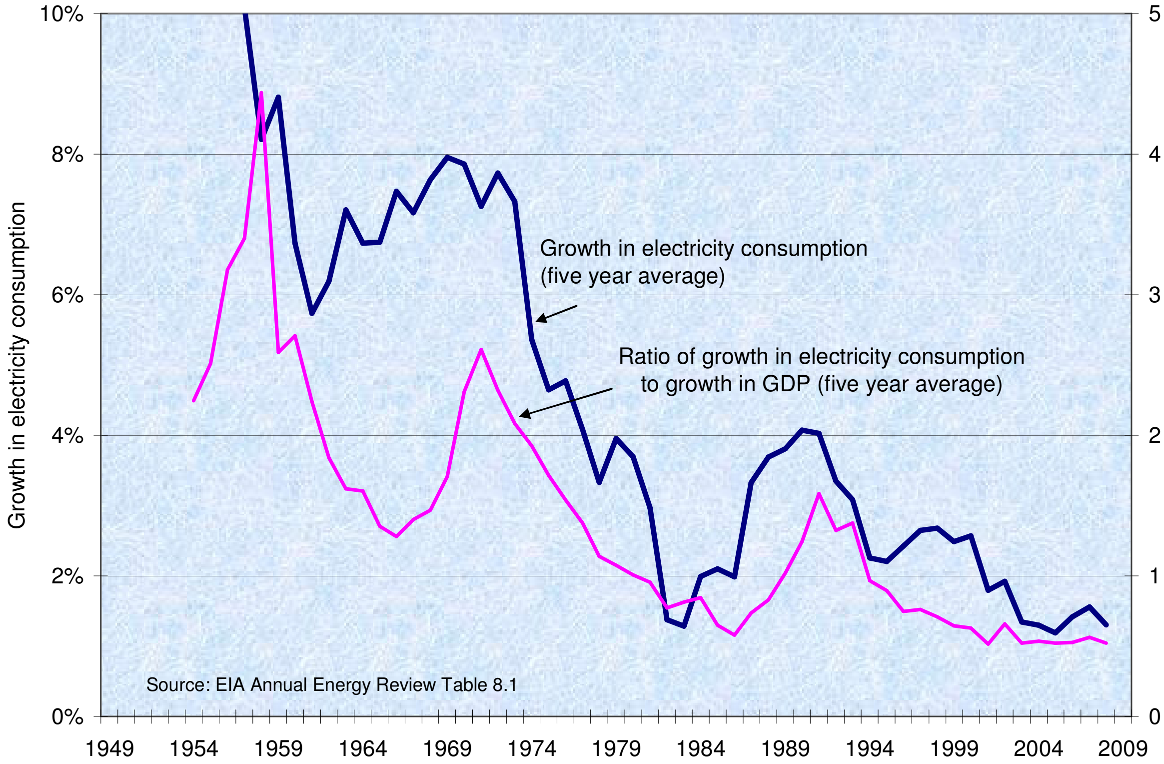
The 1-in-10 criterion, and conservative approaches to its application, apparently became widely accepted decades ago when electricity demand in the U.S. grew at a fairly steady rate and continuously required generating capacity additions. Under such circumstances, if utility resource planning was conservative and targeted high reserve margins, the excess capacity was never excess for long. In addition, the power plants built to meet incremental needs in the past required years to build, and with rapidly growing demand, the risk and potential cost of not beginning construction in time was substantial. Under these circumstances, the costs and risks favored targeting large reserve margins and building needed capacity well in advance.

Resource adequacy may have been, and remain, very conservative for additional, very human, reasons. A much higher frequency of distribution system outages has been tolerated by planners and regulators, perhaps because of the inevitability of the acts of nature or component failures that are the proximate causes of such outages. While the frequency of distribution system outages can be reduced through more aggressive vegetation management and other practices, these outages generally have proximate causes that are not under the control of utility planners or regulatory authorities. In contrast, outages due to inadequate resources seem preventable – a few more MW would have reduced or eliminated the need for firm curtailment -- and thus suggest a failure by the utility to build enough capacity in time, or by the regulatory authority to issue permits or approve cost recovery for new construction. Providing a very high level of resource adequacy may also reflect greater concern on the part of utility planners over reliability (for which they are responsible) than its cost (which is passed on to consumers). Having abundant resources arranged well in advance also makes both planning and operation of the system easier.

B. Recent changes in the circumstances of resource adequacy planning

The historical conditions that may have contributed to acceptance of highly conservative approaches to resource planning have been changing. Growth in electricity demand has become more variable and generally slowed, both in absolute terms and relative to economic growth (Figure 5). At present, peak demand growth and the need for capacity additions in the near term are more uncertain due to the downturn in the economy that began in 2008, rising electricity prices over the past few years, and recent accelerated efforts to achieve greater energy efficiency. State and federal energy policies have encouraged energy efficiency and demand response for many years, and programs have been strengthened in recent years, with many states implementing enabling legislation and setting specific targets for reductions in energy use. Increasing concerns about climate change and other environmental impacts of energy production have heightened interest in efficiency and renewable sources of energy.

Figure 5: Growth in U.S. Electricity Consumption and GDP



Source: EIA Annual Energy Review Table 8.1

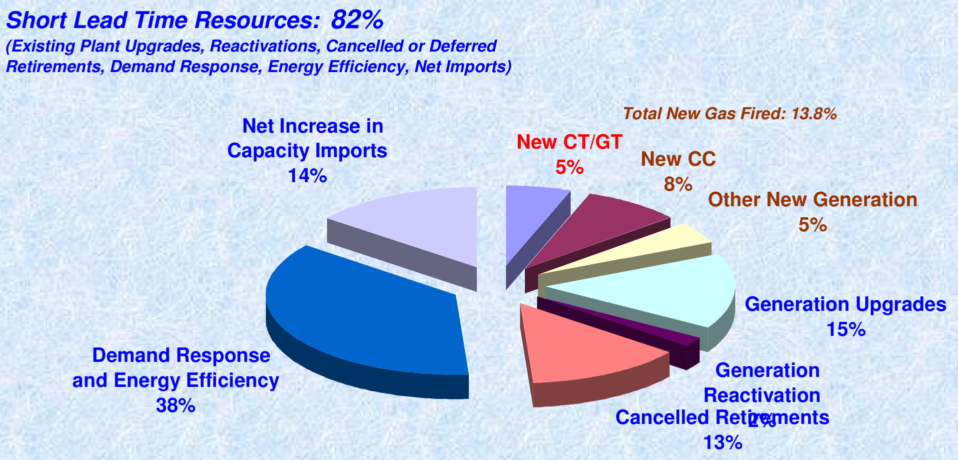
These changes are contributing to slower growth in peak loads. On the PJM system, for example, weather-normalized peak load was nearly flat from 2005 to 2008, before declining sharply in 2009 due to the recession.[31] The risk that capacity reserves built at this time may be unneeded, and may remain unneeded for years, is much higher than when the current resource adequacy practices became widely accepted decades ago.

In addition, the incremental capacity resources now being attracted to meet anticipated future needs have shorter lead times than the large fossil fuel plants that were the primary means for meeting incremental capacity needs in the past; these resources include demand response, incremental upgrades to increase the capacity or extend the life of existing plants, and deferred retirements, among other sources. Figure 6 summarizes the incremental capacity additions from the first six PJM RPM auctions based on PJM's data, and shows that short lead-time resources accounted for 82 percent of the incremental resources in these auctions, which were held to acquire capacity for peak loads in 2007 through 2012. The combination of declining peak load growth and the availability of short lead-time resources has reduced the need for and value of building large reserve margins and of acquiring them well in advance.

---

[31] PJM Load Forecast Report, January 2009, Table F-1. PJM's weather-normalized peak load grew at a 0.7%/year rate from 2005 to 2008, and less than two tenths of a percent from 2007 to 2008, before declining 1.9% in 2009.

Figure 6: Incremental Capacity Resource Additions: First Six PJM RPM Base Residual Auctions



Source: PJM, *2012/2013 RPM Base Residual Auction Results*, Tables 5 and 7; based on offered capacity expressed in installed capacity terms except DR.

C. Resource adequacy and the coming smart grid

Earlier sections of this paper showed that traditional approaches and attitudes toward resource adequacy planning have been extremely conservative, and an adjustment to these approaches would appear to have been justified, and in the consumers' interest, even based on the world of the late twentieth and early 21st century. However, the changes underway as we look forward to the second decade of the 21st century provide additional reasons to reconsider the long-standing approaches to resource adequacy.

Very broad-based efforts are underway and accelerating to implement the complex of substantial changes to the electricity industry included under the heading "smart grid." The smart grid will involve upgraded control and communications and advanced metering infrastructure ("AMI"), which will allow real-time communications to and from customers and their "smart" end-use devices.

The Energy Independence and Security Act of 2007 funded smart grid research and called for demonstration programs for smart grid elements, and the American Recovery and Reinvestment Act, signed into law in early 2009 to stimulate the economy, includes $4.3 billion for smart grid technologies. Many pilot studies have been successfully completed, and utilities across the nation are proposing to install millions of advanced meters in the next few years. While some of the elements of the smart grid are years away, many are being implemented now, and some forecasts anticipate enormous investments in the smart grid over the next several years.

These changes will accommodate substantial increases in demand response and price-responsive demand, as documented in a recent FERC Staff Report,[32] leading to future peak loads becoming much more manageable and price-responsive. While large industrial and commercial consumers have been compensated for standing ready to provide peak load reductions for a long time, smaller commercial and residential consumers will also be offered advanced metering and pricing programs that provide incentives to reduce or delay consumption when electricity is relatively scarce or expensive. Residential consumers will be able to control how their air conditioners, refrigerators, washers, dryers and other appliances respond to such signals, reflecting their individual preferences for comfort and savings. Longer term, in addition to price-driven reductions in consumption, energy storage devices, on-site generation, and plug-in hybrid electric vehicles (PHEVs) will increasingly sell power back to the grid in peak hours, reducing capacity needs.

The strong push for advanced meters, demand response and price-responsive demand reflects the expectation that substantial benefits will result from these investments. The primary benefit of demand response and price-responsive demand is the reduction in peak load and peak capacity requirements; fewer generating plants and transmission lines will be needed. While in some estimates of the benefits of demand response the reduction in capacity needs is the only quantified benefit, one estimate identified additional benefits, still finding the reduction in peak capacity needs to represent 90% of the total benefit.[33]

Realizing the enormous potential for price-responsive electricity demand requires changes to wholesale and retail pricing approaches to encourage and reward reductions in consumption at peak times. Because reducing peak loads can obviate the need to build additional power plants, large incentives to reduce demand or sell back power at such times are economically justified. Prices for energy and ancillary services (short-term reserves) in wholesale markets will increasingly be allowed to rise when reserves are low, and potentially to rise sharply if there is insufficient price-induced supply and demand response. More and more customers will face real-time pricing, Critical Peak Pricing or Critical Peak Rebate programs[34] offering substantial inducements during peak periods to reduce or delay electricity use. These change will occur in conjunction with consumer protections at the retail level (for instance, through critical peak rebate approaches, or bill protection provisions) and wholesale level (addressing the potential for exercise of market power in peak or near-peak periods).

Pilot programs have shown that these innovations can lead to substantial reductions in peak loads. The FERC Staff Report cited earlier estimates potential 2019 reductions in peak load ranging from 6% to 13% in various parts of the country even under "Expanded Business-as-

---

[32] FERC Staff Report: *A National Assessment of Demand Response Potential*, prepared by The Brattle Group, Freeman, Sullivan & Co., and Global Energy Partners LLC, June 2009.

[33] The Brattle Group, *The Power of Five Percent: How Dynamic Pricing Can Save $35 Billion in Electricity Costs*, May 16, 2007, p. 5, quantifying the benefits of a five percent demand response as including 81% reduced generation needs, 9% reduce transmission needs, and 10% energy cost savings.

[34] See, for instance, Pacific Gas and Electric Company's Critical Peak Pricing program; details are available at http://www.pge.com/includes/docs/pdfs/mybusiness/energysavingsrebates/demandresponse/cpp/dr_cpp_1858.pdf.

Usual" assumptions, with much higher reductions under "Achievable Participation" assumptions.[35]

IV. Adapting Resource Adequacy for the Smart Grid

Ultimately, with a significant fraction of electricity demand becoming price-responsive, traditional reliability-based resource adequacy planning should become unnecessary, as described further below. During what may be a long transitional period, resource adequacy approaches should be adapted to fully anticipate, accommodate, encourage, and reward peak-reducing technologies and practices in a manner that does not jeopardize reliability. There is a significant risk that if traditional resource adequacy approaches are continued, they will lead to over-procurement and excess capacity during this period, which could discourage and delay the development of the smart grid and price-responsive demand.

A. Resource adequacy in the transition to the smart grid

The next several years will likely be a transitional period during which price-responsive demand will be developing at a pace that will vary considerably from region to region, and will be difficult to predict in advance. Peak load forecasting and capacity procurement will also be complicated by uncertainty about economic growth, energy prices, and other factors that affect peak load growth.

In recent years, demand-side reductions (primarily direct-controlled load reductions by industrial and large commercial customers) have been treated as capacity resources for capacity planning purposes in some regions. Load reductions resulting from energy efficiency measures can also be treated as capacity resources in some regions.[36] However, treating demand-side reductions as capacity resources requires identification of "baseline" consumption levels, detailed measurement and verification provisions, and other administrative complexities. As the quantities grow, concerns grow about the performance and reliability of these reductions. There are also concerns about the potential for gaming to maximize payments under these programs. Peak load forecasting also becomes problematic, as load reductions treated as capacity resources (but not other reductions) should in principle be included in peak load forecasts to not double-count reductions. Over the longer term, the objective should be to de-emphasize or phase out the practice of treating demand reductions as capacity resources, which is unique to the electricity industry; ultimately, consumers should pay based on what and when they consume, as in other industries.[37] However, this requires further development of wholesale and retail pricing and

---

[35] FERC Staff Report, Figure ES-3, p. xiii.

[36] ISO New England and PJM permit qualifying demand response and energy efficiency to be offered as capacity resources satisfying capacity obligations in their respective forward capacity mechanisms.

[37] This view – that treating demand reductions as resources is a transitional measure – was reflected in a recent statement of PJM's CEO expressing PJM's vision of the development of price-responsive demand. The statement referred to demand response programs as the "second generation" of demand response improvements, with the next generation being price-responsive demand, expected to represent the majority of demand response in the future. Statement of Terry Boston, President and CEO, on behalf of the PJM Board of Managers: *Demand Response in the PJM Markets*, June 26, 2009.

other market developments. In the meanwhile, treating demand response as a resource can provide a strong incentive for further development of this valuable capability.

1. Forecasting peak loads and load reductions

Resource adequacy planning should fully anticipate the potential for peak load reductions, including the reductions induced by the potential for prices to rise to very high levels when there is risk of low reserves. A portion of the reductions may be treated as resources and recognized in supply plans, with the remainder recognized in peak load forecasts. The challenge during the transitional period is to forecast, without double-counting, all reasonably anticipated future peak load reductions and energy efficiency improvements.

Unfortunately, all types of load reduction are highly uncertain at this time, even looking out just a few years. The pace with which the various smart grid elements will be put in place will vary substantially from state to state and utility to utility. There are already large differences between states in their goals for energy efficiency and renewables, and some states and utilities will move aggressively to implement advanced metering and peak period pricing while others will choose a slower path.

While utilities and RTOs will play a large role in implementing the smart grid, they will understandably be cautious about anticipating, in their long-term peak load forecasts and capacity planning, significant impacts of unproven targets and programs regarding efficiency, demand response and price-responsive demand. RTOs especially will have little control over these developments and will not be able to ensure that programs intended to realize peak load reductions will be implemented in a timely and effective manner. Therefore, peak load forecasting and capacity planning approaches will likely continue to be conservative and to reflect a "believe it when I see it" approach to much of the potential for peak load reductions.

Econometric forecasting approaches that extrapolate past trends to predict future peak load levels generally will not accurately project new and accelerating trends, such as efficiency improvements and increasingly price-responsive demand, except with a substantial lag. Forecasting approaches that recognize specific end uses are more easily adapted to anticipate the increasing penetration and deployment of advanced meters and efficient and smart end-use devices.

Because the extent to which future peak loads become price-sensitive will vary from area to area and is largely under the control of states, utilities, and load-serving entities ("LSEs"), these entities should play an active role in forecasting the impacts of these programs and determining the capacity requirements they displace in each area. However, in some RTO regions, the RTO prepares the peak load forecasts for the entire RTO region and its sub-regions. Affording a greater role in forecasting future peak loads to states, utilities and/or LSEs could potentially reduce the risk that these forecasts will fail to reasonably anticipate future peak load reductions. In any case, policies should be clear that should emergency purchases or firm curtailments be required due to a shortage of capacity, the costs and curtailments would be allocated to those entities that are short on capacity in proportion to their shortages, to the extent this is feasible.

If utilities and RTOs continue existing, conservative approaches to forecasting and planning for future capacity needs, they will maintain adequate capacity, but they will also undermine the actual need for and value of smart grid enhancements and peak-reducing capability. If reserves continue to be planned for the "dumb peak" (reflecting only contractually committed demand response providers), there will be excess capacity and infrequent instances of low reserves and high prices, and, therefore, only weak price incentives for electricity consumers to invest in and deploy smart appliances and other peak-reducing technologies that realize the majority of their value at such times. Consumers could end up bearing both the cost of the excess capacity, and also the cost of the advanced meters and smart devices that realize little of their value due to the excess capacity conditions. In addition, the anticipated excess capacity, by reducing the need for and value of price response and energy efficiency, serves as a disincentive to achieving the targets set for their development.

2. Reserve margins and capacity requirements

During the transitional period, reserve margins and capacity requirements should be calculated taking into account the "smart peak" that results when all anticipated peak load reductions are realized. Because adequacy criteria focus on loss of load, reserve margin calculations should focus on the peak load levels that result under the most extreme scarcity pricing/low reserve conditions, with prices and incentives at the highest possible levels and leading to maximum load reductions.

As described earlier in this paper, under current circumstances, reserve margins and capacity requirements are driven by extreme peak load conditions that occur very rarely. As peak loads become more manageable, capacity requirements will decline and be targeted to high load levels likely to occur with much greater frequency (peak load reductions will tend to shift some consumption to morning and evening "shoulder" periods on the same day, flattening but also broadening the daily peak). If loads on the highest load days can be managed (as needed) down to the load levels that occur many more days each year (as illustrated in Figure 7, based on peak load data from the PJM system in 2006), much less capacity will be needed to satisfy the 1-in-10 criterion, and the least-used capacity will be needed in many more hours.

3. Capacity procurement during the transition

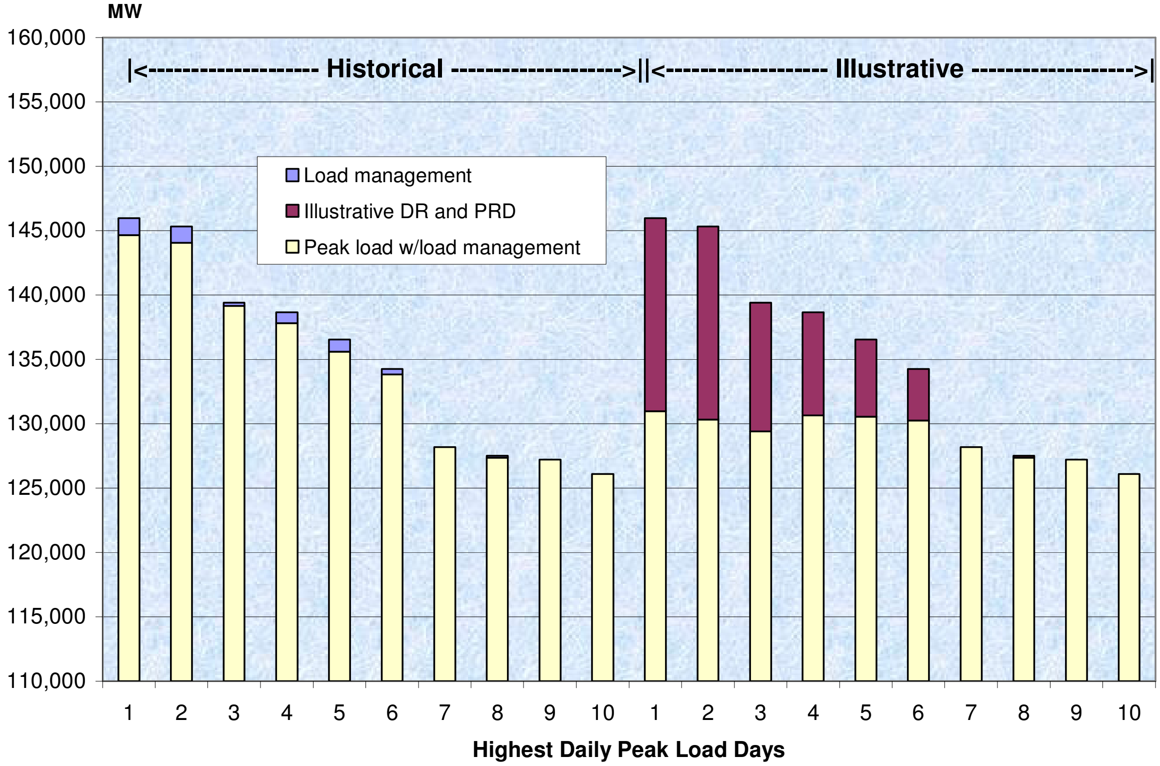
As noted above, increases in peak load and capacity requirements are presently much more uncertain than in the past, due to economic conditions and demand-side developments. However, many of the incremental resources available at this time have fairly short lead-times. These circumstances suggest that maintaining flexibility in capacity procurement is valuable and can reduce the risk and potential cost of procuring excess capacity, or paying excessive prices for capacity, years in advance.

The risk of procuring excessive amounts of capacity is greatest where capacity requirements are determined and procurement obligations are imposed years in advance (in particular, in the ISO-NE and PJM RTOs). Through its RPM capacity mechanism, PJM has already acquired over 10,000 MW of capacity for each of the four delivery years 2009 through 2013 in excess of its capacity requirement, when that requirement is updated based on an updated economic forecast.[38]

---

[38] Author's calculation, based on an economic forecast centered on Aspen Publisher's "Blue Chip Consensus" GDP forecast, October, 2009. The 10,000 MW of excess capacity includes, in addition to forecast error, the impact of RPM's sloped demand curve and a downward revision to PJM's reserve margin.

Figure 7: Peak Loads and Load Management



When substantial short lead-time resources are available, three-year advance mandatory procurement is not necessary to ensure adequate lead time to acquire needed resources for reliability. In addition to risking procurement of capacity that proves unneeded, mandatory advance procurement may also exclude or impose risks on shorter lead-time resources, because many of them may not have been identified or may not be prepared to offer their capacity so far in advance. Nor does mandatory forward procurement contribute in a significant way to the ability to attract major new power plants; such facilities require long-term contracts or other long-term revenue assurance beyond what is offered through RTO capacity mechanisms.

The risk of excess procurement associated with forward procurement obligations at this time can be reduced by providing additional flexibility in fulfilling the forward purchase obligations. A small step in this direction is accomplished by reducing the fraction of anticipated requirements that must be purchased the full three years in advance to less than 100%, in recognition of the availability of short lead-time resources, and also the uncertainty of the peak load forecast. PJM has implemented this change to its RPM mechanism. Additional flexibility for market participants to shift purchases and sales between the years-forward auctions and those closer to each delivery year, subject to limits to protect against market power, would contribute to market efficiency and reduce the risk of excess procurement and excessive capacity prices in forward markets. Allowing “virtual capacity” offers in the forward markets would provide such flexibility.

When substantial short lead-time resources are available, should it ever appear that a capacity shortage may be developing, incremental purchases of these resources can fill the gap. For example, state- and utility-level programs designed to attract additional demand-side resources could be accelerated as needed to cover an anticipated shortfall in these or other resources, or an older unit scheduled for retirement could be contracted for a few additional years.

Adapting capacity procurement rules to afford greater flexibility will reduce the risk of procuring unneeded capacity that also pre-empts short lead-time resources and peak load reductions that may be more desirable and cost-effective ways to balance supply and demand.

B. Resource adequacy on the smart grid

Longer term, as peak loads become increasingly price-sensitive and manageable, traditional resource adequacy planning approaches based on adequacy criteria will no longer be needed. Modeling loss of load risks to determine a reserve margin to satisfy a loss of load criterion will become both unnecessary and unworkable.

The common assumption that peak load is independent of supply availability will no longer hold, because prices will increasingly link peak demands (and especially the extreme peaks traditionally associated with loss of load risk) to supply conditions. When prices and incentives can reach high levels, possibly approaching VOLL, to call forth the maximum price response, modeling the circumstances under which firm curtailment could occur (the basis of LOLE studies to determine 1-in-10 reserve margins) becomes both more difficult and less meaningful. It becomes more difficult to model firm curtailment circumstances because actual peak loads become highly dependent upon system conditions. It becomes less meaningful to distinguish the circumstances under which firm curtailment might occur because, by definition, the average firm customer is close to indifferent between paying prices close to VOLL or being curtailed.

When peak loads become highly manageable, the “symptom” of inadequate capacity will no longer be an unacceptable frequency of anticipated involuntary firm curtailment, as assumed under the traditional approaches to resource adequacy based on LOLE criteria. Instead, the symptom of inadequate capacity will be too many hours with high prices and substantial voluntary customer reductions, imposing a cost on these customers that exceeds the incremental cost of additional capacity. The optimal level of capacity will balance the cost of additional capacity and the benefit of avoiding price-induced demand reductions, as a need to involuntarily curtail firm loads due to inadequate resources becomes increasingly unlikely. Put differently, supply additions will result from the interplay of supply and demand, as for other goods and services, rather than administrative reliability rules and obligations. Ultimately, the 1-in-10 criterion that has been so conservative to the present time, to the extent it will still be possible to meaningfully apply it, will suggest levels of capacity less than the optimal amounts, and less than the amounts that will be provided under market incentives.

As markets and prices rather than reliability rules begin to determine capacity levels, the RTOs’ capacity mechanisms can be phased out, as such payments will no longer be needed to achieve acceptable levels of reliability. The shift away from revenue recovery through capacity payments will also help achieve goals for attracting renewable resources, whose capacity values are typically deeply discounted.

C. The future should hold more instances with (at least a chance of) low reserves

In the past and at present, with limited ability to manage peak loads, providing a high level of reliability requires planning a capacity margin over extreme peak load levels that are unlikely to occur. As a result, there is nearly always abundant capacity and operating reserves rarely fall below desired levels. As peak loads become more manageable and price-driven, resources should be planned assuming this capability will actually be used. Even during the transitional period, the planning outlook should include more frequent short periods of low reserves and rising prices -- or at least the potential for these conditions, unless there is sufficient demand and supply response. If there is sufficient response, price spikes will not occur, or will be brief and muted.

Activation of demand-side reductions should not be considered indicative of a failure to plan and build adequate resources. These actions become part of the plan, to be expected, up to a point. In addition, instances of high prices and critical peak rebates, when consumers realize the value of their investments in smart meters and devices, will have a positive impact on sales and deployment of such devices.

V. Summary

To summarize the main points of this paper:

1. Planning for adequate electric generating capacity has always been very conservative, erring on the side of too much rather than too little capacity. The criterion used – "one day in ten years" – appears to be roughly an order of magnitude more stringent than justified based on a balancing of incremental cost and benefit. Yet in practice, planners and policy-makers have achieved even higher levels of capacity and resource adequacy. Highly conservative resource planning made some sense in the past when incremental needs had to be met with new power plants that took years to build, and any excess capacity was soon needed due to rapidly rising electric demand.

2. Changes in the electricity industry are changing the need for and benefits of planning substantial capacity margins to meet anticipated peak loads, and of contracting the capacity years in advance. Electricity demand growth has become more variable and generally slowed, both in absolute terms and relative to economic growth. At present, peak demand growth and the need for capacity additions in the coming years is more uncertain due to the downturn in the economy that began in 2008, higher electricity prices, and renewed efforts to achieve greater energy efficiency and demand response. The risk that capacity built at this time may be unneeded and remain unneeded for years is higher, and the potential cost of excess capacity is much higher, than when the current conservative approaches to resource adequacy and the 1-in-10 criterion became widely accepted decades ago. In addition, many of the incremental resources now being attracted to these markets have much shorter lead times; the availability of short lead-time resources reduces the need for and value of building large reserve margins and obtaining commitments well in advance, because additional resources can be acquired relatively quickly if the need unexpectedly arises.

3. The coming "smart grid" will bring advanced metering, time of use pricing and its variants, and smart, price-responsive appliances, among other innovations. These changes will make peak

loads increasingly manageable and price-sensitive and reduce the risk of curtailments when reserves are low. However, if current approaches to resource adequacy are continued, the realization of “smart grid” demand-side technologies and price-responsive demand may be delayed, and their value reduced. High capacity margins that fail to anticipate an increasingly active demand side, and the resulting infrequent instances of low reserves and high peak period prices and incentives, result in little actual need or incentive for smart appliances or energy storage technologies. If policies encouraging or requiring large capacity margins are continued, consumers may be saddled with both the cost of the excess generating capacity and also the cost of the advanced meters and smart devices that will realize only a fraction of their potential.

4. To encourage and accommodate demand-side resources that are smarter and more price-sensitive, resource adequacy approaches should be adapted. Load forecasts should anticipate all types of price-responsive demand. Under RTOs, a greater role in forecasting of demand-side reductions should rest with entities closer to the customer – the states, utilities and load-serving entities -- who are also responsible for realizing the programs and meeting capacity requirements. This will also better accommodate the differences from state-to-state and utility-to-utility in the potential for advanced metering, critical peak pricing, and price-responsive demand, and the pace at which the potential will be realized.

5. The capacity procurement process should become more flexible with less emphasis on acquiring large capacity margins far in advance. With uncertain peak load growth and a variety of available short lead-time resources, procurement can shift closer to the time of apparent need when the need is clearer. Where capacity obligations are imposed years in advance, additional flexibility should be provided in the timing of fulfillment of such obligations.

6. As peak loads become more manageable, capacity requirements will decline and target smart peak load levels that occur more frequently, rather than extreme, infrequent “dumb” peaks. The optimal level of capacity will increasingly reflect a balancing of the cost of building capacity against the cost of frequent, voluntary demand reductions. The “one day in ten years” criterion, focused on involuntary firm curtailment, will become difficult to apply and less meaningful, and ultimately will not be needed to maintain reliability.

7. More frequent instances of low reserves, and higher energy and ancillary services prices at such times, will increase the earnings from these markets to all resources, allowing capacity payments to be reduced and eventually eliminated. The current RTO markets that emphasize capacity payments also discourage most renewable sources of supply, whose capacity values are heavily discounted; phasing out capacity payments will contribute to encouraging renewable resources.