GENERATION AND TRANSMISSION ADEQUACY IN A RESTRUCTURING U.S. ELECTRICITY INDUSTRY

ERIC HIRST, BRENDA KIRBY, and STAN HADLEY
Consulting in Electric-Industry Restructuring
106 Capital Circle
Oak Ridge, Tennessee 37830

June 1999

Prepared for
Edison Electric Institute
701 Pennsylvania Avenue NW
Washington, DC 20004-2696
CONTENTS

SUMMARY .............................................................. v
LIST OF ACRONYMS .................................................... vii
1. INTRODUCTION ....................................................... 1
   BACKGROUND .................................................... 2
   DATA AND PROJECTIONS .......................................... 4
2. GENERATION ADEQUACY ............................................. 7
   CONCEPTS ........................................................ 7
   DISCUSSION ..................................................... 10
   ISO APPROACHES ................................................ 14
3. QUANTITATIVE ANALYSIS OF GENERATION ADEQUACY ............... 17
   OAK RIDGE COMPETITIVE ELECTRICITY DISPATCH MODEL ........ 17
   STRUCTURE OF ANALYSIS ........................................ 18
   RESULTS ........................................................ 19
   DISCUSSION ..................................................... 22
4. TRANSMISSION ADEQUACY .......................................... 23
   CONCEPTS ........................................................ 23
   LESSONS FROM THE GAS INDUSTRY .............................. 26
   LOCATIONAL PRICING AND CONGESTION CONTRACTS .......... 28
   TCCs AND TRANSMISSION INVESTMENT ........................... 31
   TRANSMISSION-PRICING PRINCIPLES ............................. 33
   TRANSMISSION-EXPANSION APPROACHES .......................... 34
   DISCUSSION ..................................................... 41
5. SUMMARY AND CONCLUSIONS ....................................... 45
ACKNOWLEDGMENTS .................................................. 46
REFERENCES .......................................................... 47
SUMMARY

Historically, decisions on the amounts, locations, types, and timing of investments in new generation and transmission have been made by vertically integrated utilities with approval from state public utility commissions. As the U.S. electricity industry is restructured, these decisions are being fragmented and dispersed among a variety of entities (Table ES-1).

As generation is deregulated and becomes increasingly competitive, decisions on whether to build new generators and to retire, maintain, or repower existing units will increasingly be made by unregulated for-profit entities. These decisions will be based largely on investor assessments of future profitability and only secondarily on regional reliability requirements. In addition, some customers will choose to face real-time (spot) prices and will respond to the occasionally very high prices by reducing electricity use at those times. During the transition from a vertically integrated, regulated industry to a deintegrated, competitive industry, government regulators and system operators may continue to impose minimum-installed-capacity requirements on load-serving entities. As the industry gains experience with customer responses to real-time pricing, these requirements will likely disappear.

Transmission will continue to be regulated, although regulatory approvals will increasingly come from the Federal Energy Regulatory Commission rather than from state regulators (which, however, will continue to oversee facility siting and environmental compliance). Transmission systems that were originally designed to serve native load will be required to carry significantly increased bulk-power transactions with only limited additions to the grid. Decisions on new transmission will be complicated by strong network interactions and economies of scale and scope. Addition or modification of any element in a transmission grid is likely to affect flows (including losses and congestion) in other elements, leading to large externalities. In addition, overbuilding a transmission element is usually much cheaper per megawatt of capacity than sizing it for immediate needs and upgrading it later. The planning processes being developed by regional transmission operators are slow because they require cooperation between the system operator and transmission owners and are conditional in part on the decisions of investors on the locations and timing of new generation.

Because of these factors, as well as the monopoly nature of transmission, markets may play only a modest role in affecting the type, timing, and scope of investments in new transmission facilities. Transmission-congestion pricing will, however, provide important information on how markets value transmission expansion (as well as the location of new generators); in addition, congestion pricing can have large effects on transmission operations.
### Table ES-1. Key issues related to generation and transmission adequacy

<table>
<thead>
<tr>
<th>Planning coordination</th>
<th>Past</th>
<th>Possible future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generation and transmission planned together by vertically integrated utilities</td>
<td>Competitive generation and regulated transmission planned separately; coordination more difficult to achieve</td>
</tr>
</tbody>
</table>

| Level of adequacy     | Determined by utility with (often tacit) approval from state regulator | Generation adequacy determined by market participants (generators, suppliers, and customers); transmission adequacy determined through public process with state or FERC (or both?) approval |

| Pricing               | Embedded-cost pricing had little effect on either generation or transmission adequacy | Real-time energy pricing for some retail load will affect generation adequacy; congestion pricing will guide transmission investments and locations of new generation |

| Lead times for planning and construction | Long lead times for new generation; shorter lead times for transmission | Short lead times for some (e.g., gas-fired) generation; long lead times for new transmission |

| Roles of markets and regulation | State regulation and central planning dominated adequacy decisions | Markets dominate generation-adequacy decisions and affect transmission-adequacy decisions; regulatory authority may shift from states to FERC |

Coordination between transmission and generation planning will likely suffer in a competitive, deintegrated electricity industry. For competitive reasons, investors will not reveal their generation-expansion decisions any sooner than siting regulations require. Also, the time to construct many generating units is now often shorter than the time to build a new transmission line, further complicating transmission planning. Siting new transmission facilities may get more difficult as population increases and pressures mount for protection of open spaces. On the other hand, creation of large regional transmission entities should facilitate transmission planning and expansion processes.

Some of the issues discussed here are enduring, but many are transitional. Transmission adequacy may be a long-term problem, while generation adequacy may be more of a transitional problem. Unfortunately, the transition from a highly regulated, vertically integrated industry to one dominated by competition is likely to take another five to ten years. During the transition period, all market participants are faced with the current “hybrid from hell.”
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIA</td>
<td>Energy Information Administration</td>
</tr>
<tr>
<td>ERCOT</td>
<td>Electric Reliability Council of Texas</td>
</tr>
<tr>
<td>FERC</td>
<td>U.S. Federal Energy Regulatory Commission</td>
</tr>
<tr>
<td>ISO</td>
<td>Independent system operator</td>
</tr>
<tr>
<td>LOLP</td>
<td>Loss-of-load probability</td>
</tr>
<tr>
<td>LSE</td>
<td>Load-serving entity</td>
</tr>
<tr>
<td>NEPOOL</td>
<td>New England Power Pool</td>
</tr>
<tr>
<td>NERC</td>
<td>North American Electric Reliability Council</td>
</tr>
<tr>
<td>ORCED</td>
<td>Oak Ridge Competitive Electricity Dispatch model</td>
</tr>
<tr>
<td>O&amp;M</td>
<td>Operations and maintenance</td>
</tr>
<tr>
<td>PJM</td>
<td>Pennsylvania-New Jersey-Maryland Interconnection</td>
</tr>
<tr>
<td>PUC</td>
<td>Public utilities commission</td>
</tr>
<tr>
<td>PX</td>
<td>Power exchange</td>
</tr>
<tr>
<td>RTO</td>
<td>Regional transmission organization</td>
</tr>
<tr>
<td>TCC</td>
<td>Transmission-congestion contract</td>
</tr>
<tr>
<td>Transco</td>
<td>Transmission-owning and -operating entity</td>
</tr>
<tr>
<td>VOLL</td>
<td>Value of lost load</td>
</tr>
</tbody>
</table>
Restructuring the electricity industry in general, and the bulk-power sectors in particular, calls into question the entire concept of adequacy. What is adequacy? Is it purely a reliability concept, or does it also have commercial significance? Are generation and transmission resources complements, substitutes, or both? Is adequacy even a relevant term for a restructured electricity industry?

Because of the dramatic changes under way in the ownership, operation, and structure of bulk-power systems and markets, the Edison Electric Institute asked us to explore possible changes in generation and transmission adequacy. This project addresses the conceptual issues raised above, generation adequacy, transmission adequacy, and the integration of transmission and generation planning and expansion. These questions are important and difficult because the United States is unbundling the traditional, vertically integrated utilities that, historically, managed generation and transmission adequacy within a single organizational entity. These functions are increasingly being shared among at least three entities: competitive generation, regulated transmission, and regulated system operations.

Resolving these issues is difficult for several reasons: (1) generation is likely to become increasingly competitive and deregulated while transmission remains regulated; (2) transmission operations are likely to be combined into large, independent regional organizations, the structure of which is far from clear; (3) because of these differences, decisions on generation adequacy might be left to competitive markets while decisions on transmission adequacy continue to be made, at least in part, by regulators and central planners; (4) adequacy and security are both complements and substitutes; and (5) generation and transmission are both complements and substitutes.

The bulk of our effort in this project involved collection and interpretation of written and oral materials on these topics. We analyzed historical data and projections from the North American Electric Reliability Council (NERC) and the Energy Information Administration (EIA); read articles from The Electricity Journal, Public Utilities Fortnightly, Federal Energy Regulatory Commission (FERC) filings and orders, independent system operator (ISO) tariffs and reports, the IEEE Power Engineering Society, and a wealth of materials posted on the Internet. We interviewed people from the various sectors of the bulk-power industry on these topics. We used the Oak Ridge Competitive Electricity Dispatch model (ORCED) to analyze the effects on consumer costs and prices, producer profits, and generation mix of required capacity reserve margins vs reliance on markets to determine the retirement of existing generating units and the construction of new ones.
The rest of this chapter provides background on the concept and definition of adequacy, and presents historical data and projections on transmission and generation investments and capacity. Chapters 2 and 3 focus on generation adequacy, while Chapter 4 deals with transmission adequacy. Chapter 5 presents our summary and conclusions.

BACKGROUND

NERC defines reliability as “the degree to which the performance of the elements of [the electrical] system results in power being delivered to consumers within accepted standards and in the amount desired.” NERC’s definition of reliability encompasses two concepts, adequacy and security. Adequacy, the subject of this report, is defined as “the ability of the system to supply the aggregate electric power and energy requirements of the consumers at all times.” Security is “the ability of the system to withstand sudden disturbances.”

In plain language, adequacy deals with planning and investment, and security deals with short-term operations. Adequacy implies that there are sufficient generation and transmission resources available to meet projected needs plus reserves for contingencies. Security implies that the system will remain intact even after outages or other equipment failures occur. Although adequacy is a reliability concept, it has strong commercial implications; the same is true of security. Indeed, although we might like to pretend otherwise, bulk-power reliability and commerce are strongly interdependent.

Obviously, adequacy and security are complements. Without system security, the output of the generation resources, no matter how abundant, cannot be delivered to customers. Correspondingly, a high degree of security is of little value if there are insufficient generation and transmission resources to meet customer needs.

The substitution nature of adequacy and security is not so obvious. What we mean here is that more of one can make up for less of the other. For example, an abundance of generation and transmission resources makes it easier to maintain a high degree of security (i.e., reduces the need for emergency actions). That is, system operators can manage the system in real time with less data and fewer analytical tools if there are ample generation resources and redundant transmission facilities. Similarly, high-quality system operation can extract more output from a system that might otherwise be considered underbuilt. For example, the near-real-time collection and analysis of data on the current and projected states of the transmission system can allow system operators to run the system closer to its limits than would less data collection and analysis.*

*Voltage reductions, as well as automatic underfrequency and undervoltage load shedding, could be considered a form of security that substitutes for resource adequacy. We do not include such customer-interruption and degradation-of-service measures in our discussion of substitution between adequacy and security.
The June 1998 price spikes in the Midwest illustrate this substitution effect well, in both the planning and operational senses (Terhune 1998). Early in 1998, it was clear that a substantial amount of nuclear generation would not be operational during the coming summer. As a consequence, both the Mid-American Interconnected Network and the NERC Reliability Assessment Subcommittee performed extensive studies of potential summer conditions. In addition, the utilities in the region expanded their capability to import power and to deliver power reliably within the region. These steps included a new operating guide for a 345/161-kV transformer, addition of a breaker to a 345-kV line, installation of a series reactor on a transformer, replacement of a 345/161-kV transformer, reactivation of mothballed generating units, installation of new circuit breakers, and reconfiguration of a 345-kV bus tie. Extra preparedness measures included line inspections and clearance of vegetation for high-temperature operation, cleaning of critical substation equipment, development of short-time ratings and operating guides, and inspection and addition of capacitor banks to support voltage stability. Generation preparations during the Spring of 1998 included purchase of additional replacement power, expansion of interruptible load contracts and other demand-side management programs, and aggressive preventive-maintenance efforts to maintain generator availability. These actions to improve security allowed the system operators to serve higher demands reliably.

A restructured, competitive electricity industry will reduce the integration between generation and transmission planning. The Energy Modeling Forum (1998) emphasized the close coupling between generation and transmission for both planning and operations:

… decisions about generating and transmitting power are closely intertwined. The daily operation of the transmission system depends critically upon where and when to generate power. Longer-run decisions about investing in generation or loads are closely linked to those concerned with expanding the transmission system. The existence of these interrelationships, or complementarities, between functions presents opportunities to operate and expand both systems more efficiently or at a lower cost when done jointly rather than separately. A fundamental issue in restructuring concerns how to decentralize decisions about generation and loads and still acknowledge the complementarities between generation and transmission.

Historically, it took up to a decade or more for utilities to plan, gain regulatory approval for, and build new generating units and new transmission lines. Public and environmental concerns have, for at least a decade, complicated and lengthened the transmission-planning and -expansion process. The transmission-planning processes being established by ISOs are inherently slow because (1) they are serial (i.e., transmission is planned only after new generation facilities are announced), (2) they encourage participation by all stakeholders, and (3) they require cooperation and agreement between the ISO and the transmission owners. This slowdown in transmission construction is occurring at the same time that new gas-fired
technologies and competitive generation markets are shortening the time it takes to plan and build new generation.

DATA AND PROJECTIONS

The U.S. electricity industry is currently in a very awkward position—half regulated and half competitive. Many utilities are understandably reluctant to make investments until the rules and the separation between competitive and regulated activities are clear. In addition, the pressures to cut costs and lower prices is forcing operations and maintenance (O&M) budgets down.

Figure 1, based on data reported annually by U.S. investor-owned utilities on FERC Form 1 and compiled by EIA, shows that transmission investment relative to total energy production declined slowly between 1990 and 1996. (EIA did not publish comparable national data for 1997.) Transmission maintenance relative to energy production declined more rapidly during this time. It is unclear whether these declines represent a degradation in transmission-system reliability or improved productivity.

Data reported to NERC (1998b) are consistent with the EIA data. The NERC data show a 16% decline in miles of transmission lines per MW of summer demand from 280 in 1989 to 236 in 1997 (Fig. 2). Transmission capacity is expected to decline further to 205 miles/MW in 2007.

Fig. 1. Trends in annual transmission maintenance expenses and investment for U.S. investor-owned utilities, normalized by annual electricity production.
With respect to generation adequacy, Fig. 3 shows utility forecasts, as reported to the regional reliability councils and NERC (1998a), of generation-capacity margins from 1990 through 1998 and projections through 2007. Nationwide, reserve margins declined from 22% in 1990 to 16% in 1997 and are expected to decline further to 10% in 2007. The regional trends show an especially precipitous drop in reserve margin in the Electric Reliability Council of Texas (ERCOT). As NERC notes, the projections for the last five years (2003 to 2007) are highly uncertain. This uncertainty occurs because the owners of merchant plants often do not reveal their plans early and because new generating units can often be constructed in only a few years (reducing the need for long-term projections of generating capacity).

These data and projections paint a consistent picture. For at least the past several years, both generation and transmission adequacy, appropriately normalized, have declined. Utility reports to NERC suggest that these trends will continue for the next decade. Indeed, the latest NERC (1998a) reliability assessment is more pessimistic than earlier ones, primarily because of the restructuring changes under way in the industry. This pessimism relates to a reluctance on the part of utilities to build new generation and transmission because of uncertainties about cost recovery for such investments, loss of integration between generation and transmission planning, reluctance of independent power producers to reveal their generation plans much in advance of actual construction, possible double-counting of some generating capacity as more suppliers rely on purchases from other entities, and uncertainty over the extent to which
demand-side responses will reduce the need for new generation. On the other hand, independent power producers plan to build 69,000 MW of new capacity.

NERC notes that “… the level of uncertainty has increased tremendously. Purchases from undisclosed resources and the reluctance of generation developers to disclose plans for future capacity additions are making modeling for long-term transmission analysis virtually impossible. … “Very few bulk transmission additions [are] planned. Only 6,558 miles of new transmission (230 kV and above) are planned throughout North America over the next ten years. This is significantly lower than the additions that had been planned five years ago.”

In summary, electric utilities are reluctant to make major capital investments given the uncertainties they face about the future structure of the electricity industry and their opportunities to earn adequate returns on these investments in regulated or competitive markets. Unregulated companies that build merchant plants are reluctant to reveal their plans any sooner than the regulatory permitting process requires. As a consequence, transmission planners have insufficient data about the locations, types, and sizes of new generating units to design appropriate transmission-expansion projects. Government agencies and reliability entities face growing difficulties in obtaining consistent and complete data on existing and planned generation and transmission facilities, a consequence of the increasing competitiveness of the industry and the growing diversity of entities that own and operate such facilities. Finally, the time to construct new transmission facilities has increased to the point that it often takes less time to build a gas-fired generating unit than a transmission line.
CHAPTER 2

GENERATION ADEQUACY

CONCEPTS

Historically, utilities maintained “extra” generating resources for short- and long-term purposes (Exhibit 1); this report focuses on long-term reserves, often called planning reserves, and does not deal with operating reserves. At least two mechanisms can be used to maintain generation adequacy:

- Rely on markets, the interactions of consumers and suppliers acting through the mechanism of volatile spot prices, to decide what types of generation to build and when and how much electricity to consume when. California adopted this approach.

- Rely on the traditional system of having a central agency (e.g., the ISO or state regulator) specify an appropriate minimum reserve margin based on estimates of the value of lost load (VOLL) and other factors (e.g., forced and planned outage rates for different types of generating units). This reserve margin is then imposed on all load-serving entities (LSEs). The three Northeastern ISOs (PJM, New York, and New England), all of which developed from traditional tight power pools, use this approach.

The United Kingdom uses a third system. There, the National Grid Company calculates, on a day-ahead basis, the expected loss-of-load probability (LOLP) for each 30-minute period. This LOLP is then multiplied by the assumed VOLL of about $4/kWh to develop a capacity charge, which is added to the system marginal price; see Exhibit 2. This approach has received little attention in the United States, perhaps because the capacity charge is too easy to manipulate for companies that own large amounts of generation. They can do so primarily by declaring units unavailable in the day-ahead market and then redeclaring them available in real time to collect the high capacity charge caused by the unavailability declarations. In addition, the day-to-day and seasonal volatility in this capacity charge may make it a poor mechanism to encourage investors to build new generating capacity.

Thus, the key issue on generation adequacy is whether (1) competitive generation markets for capacity and energy will be sufficient to maintain societally desirable levels of reliability or (2) government regulators and central planners (e.g., ISOs or Transcos) will need

---

*This system of specifying a minimum planning reserve is quite different from the integrated resource planning process that many states required of their electric utilities. While IRP dealt with the technologies, fuels, and costs of generating capacity, minimum reserve margins determine only how much capacity must be installed.
to impose mandatory minimum-reserve obligations on LSEs to ensure that customers are not involuntarily interrupted from their electricity supplies.

These two options should produce different outcomes in:

- hourly energy prices, with reliance on real-time markets likely to yield lower average prices and costs but greater price volatility;
- customer load shapes, with reliance on real-time markets likely to yield higher load factors; and
- generation portfolios, with reliance on real-time markets likely to yield relatively more baseload capacity.

Mandating minimum planning reserves may necessitate the creation of markets for installed capacity, which would be traded among the LSEs in a region. The associated capacity charges would, unavoidably, distort the energy markets. Proponents of market-based methods to determine the amount of installed generating capacity worry that government regulators will not permit the kinds of price volatility required to signal markets on the need for new generation or demand curtailment.
Exhibit 2. How the U.K. System Addresses Generation Adequacy

In the U.K., the price each online generator receives each half hour is the sum of two components, the second of which is called the capacity charge:

\[
\text{Pool purchase price} = \text{SMP} + \text{LOLP} \times (\text{VOLL} - \text{SMP}),
\]

where SMP is system marginal price and is equal to the highest bid accepted for that half-hour period, LOLP is loss-of-load probability as calculated the day before by the system operator, and VOLL is value of lost load (administratively set equal to about $4/kWh). As shown in the figure below, between January 1996 and August 1997, the capacity charge averaged 15% of the system marginal price (Newberry 1998).

Wolak and Patrick (1997) note that “the strategic declaration of [generator] availability [is] a very attractive way … to obtain large values of the day-ahead spot price.” The nonlinear relationship between the expected reserve margin and the LOLP yields large benefits from strategically withholding capacity to obtain a small reserve margin and a high LOLP and, therefore, a high capacity-charge payment.

The capacity-charge term was added to provide market signals concerning investment in new generating capacity. But the evidence to date suggests that this administratively determined factor is a source of market power rather than a useful economic incentive to build new generation.

The U.K. capacity charge added about 15% to the system marginal price (SMP) in 1996 and 1997.
DISCUSSION

Our review of the literature as well as our discussions with several market participants yielded surprising agreement. Almost everything we read and everyone we spoke with believes that—in the long run—generation adequacy will be left to markets with little involvement by government regulators. To do otherwise, most people recognize, would interfere with the workings of what are supposed to be competitive energy markets. That is, the energy and capacity markets are closely coupled.

On the other hand, most people agreed that we may need a multiyear transition period while suppliers and, especially, retail customers learn how to respond appropriately to rapidly changing (e.g., hourly) electricity prices. (We first have to permit retail customers to face these time-varying prices; in most parts of the country, customers still face prices based on embedded costs that are largely time invariant.) During this transition period, prudence may require maintenance of mandated planning-reserve margins.

Proponents of market-based decisions on generation retirements and expansions worry, however, that electricity price spikes, such as occurred in the Midwest in June 1998, will bring forth inappropriate government price controls.* According to Lapson et al. (1998):

Market fluctuations heighten regulatory risk. The jury is still out [as of October 1998] on whether policy markets (legislators and regulators—elected and appointed officials) and the public can tolerate price fluctuations in the energy market. After the [June 1998 Midwest] price spike, industrial consumers, utilities, legislators, and others called for price caps or price regulation to limit prices on the upside. (No consumers or legislators have clamored for price floors to limit producers’ losses during shoulder seasons when prices are microscopic.) So far, FERC and the Congress have resisted the call for price caps. However, in the future, additional price anomalies, even for brief periods, will reduce regulators’ and politicians’ enthusiasm for a competitive electricity commodity market.

In support of the market option, Michaels and Ellig (1998) note that:

Price spikes … provide market participants with important information needed for trading and capacity investment decisions. Price increases signal price-sensitive customers that it is time to conserve, and they tell producers that it may be time to expand capacity. Price increases also give producers and consumers incentives to change their behavior in ways that mitigate severe spikes; producers can profit by investing in new capacity, and consumers can make

---

*Prices spiked, in part, because almost all retail customers paid only traditional, embedded-cost rates and did not face these very high wholesale prices.
themselves better off by reducing peak period demand. … It is true that only some customers need to moderate their usage to reduce peak prices for everyone. But in the absence of a competitive market, we have no way of knowing which customers are most willing to do this.

Figure 4 schematically illustrates the supply/demand balances with and without an explicit installed-capacity requirement. The dashed line that slopes up to the left represents consumer demand, and the stairstep line that slopes up to the right represents generating capacity. With a reserve-margin requirement of 11,500 MW, supply and demand equilibrate at a price equal to the variable cost (fuel plus variable O&M) of the last (marginal) unit online at that time. If, however, there is no required reserve margin and market forces yield only 10,500 MW of available capacity, the price of electricity will rise above the variable cost of the last unit online when unconstrained demand exceeds 10,500 MW. The amount of price increase (the pure capacity price in Fig. 4) is a function of the demand elasticity for electricity. The more responsive customer demand is to changing electricity prices, the smaller this capacity price will be. Rose (1997) writes that “This premium [pure capacity price] emerges in peak demand hours in which the chance of a shortage of generation capacity becomes significant.”

Fig. 4. With an installed-capacity requirement of 11,500 MW, supply and demand balance at a price of 3.0¢/kWh. With no capacity requirement and only 10,500 MW online, unconstrained demand would exceed supply, and prices would rise to 3.4¢/kWh.
This example makes two points. First, even if there is “insufficient” capacity from an engineering perspective, price-responsive demand and supply will equilibrate, and the bulk-power system will not crash. This equilibrium occurs because some customers would rather forego some consumption than pay the high price associated with this situation. Second, at times of high demand, spot prices will be higher if there is no required planning-reserve margin. In other words, specifying a minimum amount of installed generating capacity will suppress spot prices at certain times. Economists argue that this suppression of a valuable price signal will undercut energy and capacity markets.

Requiring a minimum reserve margin creates two markets (installed capacity and energy) with no assurance that they will be in equilibrium with each other (Graves et al. 1998). This requirement will suppress energy prices and undercut demand-side participation in reliability. Requiring minimum reserve margins and the associated two-part pricing for capacity and energy “will undermine the benefits of power industry restructuring.” On the other hand, energy-only markets “will induce efficient capacity planning—which has been the real problem in the past (not inefficient dispatch) and which is where the real opportunities for future efficiency gains lie. It will also encourage demand-side participation in peaking reserves, and forward contracting for risk protection … .”

Graves et al. (1998) note several problems with the traditional engineering approach to maintaining generation adequacy, including:

- Setting fixed capacity requirements to deal with what is inherently a very uncertain situation. The uncertainties deal with the timing, extent, and duration of forced outages and with the tremendous variation among customers in their value of lost load.

- Static demand curves with zero price elasticity in spite of the evidence from real-time pricing programs that customers differ substantially in their willingness and ability to respond to changing electricity prices.

- The assumption that forced outages occur in a completely random fashion. In competitive markets, generation owners will work hard to assure that their plants are available during high-price periods.*

- With customer choice, customers may want to choose their level of reliability, not have it specified for them by a central authority.

Henney (1998) is similarly critical of mandating minimum-reserve requirements:

This [PJM] proposal [for installed-capacity requirements] artificially separates the market into a capacity market and an energy market. Yet that is not what

*Seiple (1999) concludes that “The move to an asset-based management philosophy will induce companies to improve forced outage rates and reactivate assets that can generate additional profits, thus increasing the overall supply.”
customers generally buy: they buy kilowatt hours at different times of the day and year. … splitting capacity and energy is a carryover from the regulatory revenue collection practices of monopolies and from the traditional approach to ensuring generation reliability by physical capacity planning. Fundamentally, this provision subordinates the design of a trading market to a function that the PJM-OI has wearing another hat, namely, a responsibility for keeping the lights on by ensuring generation reliability. [This PJM approach:]

- Provides incorrect price signals about the type of plant that should be built, with a bias in favor of a peaking plant.
- Is gameable in allowing a party with a very cheap but runnable unit (e.g., a third-hand areoderivative gas turbine) access to cheap energy produced by more capital-intensive but fuel-efficient units.
- Offers a barrier to new entrants by increasing complexity.

The objective of achieving generation reliability should be to maximize the use of market mechanisms. To the extent that is not possible, the aim should be to provide minimally intrusive mechanisms for ensuring reliable supply … .

Loehr (1998) offers an opposing, cautionary view of generation adequacy:

The argument in favor of deregulation is that investors will build new units in response to the demands of the market. Perhaps so. But there are some major concerns. For most electric power systems in the United States, the actual load exceeds 90% of the peak load only 1 to 2% of the time. In the past, utilities had an “obligation to serve” all of the load all of the time; even the last 10%. This was part of the regulatory compact. Thus they planned, built and operated as much generation as was required by the peak load. But today there is a real question as to whether, in an industry driven by competition and the marketplace, investors will be willing to commit financial resources to supply customer load which will be realized only a few hours a year. As far as actual or potential generation owners are concerned, this is a basic question of price and price signal.

Loehr appears to suggest that, in a competitive electricity industry, generation owners cannot earn a profit building plants that operate only 1 to 2% of the time; given a choice, they would not build such plants. Therefore, society needs to make them do so through minimum-reserve requirements. Society then requires all electricity consumers, regardless of how highly they value electricity consumption at times of tight supplies, to pay for this capacity. Loehr appears not to consider the possibility that these plants will be unprofitable because consumers would rather reduce consumption than pay such high prices. The economists argue that this “extra” generating capacity should be built only if customers are willing to pay the very high spot prices associated with the very infrequent use of these units; otherwise, enough customers
will reduce their demand sufficiently to yield a supply/demand balance at a lower level of generating resources and lower peak-period prices.

Jaffe and Felder (1996) believe that mandated capacity requirements are needed because such capacity benefits society at large, not just the owners of such capacity (what the economists call positive externalities). Such societal benefits are especially large for electricity because of its pivotal role in modern society, the real-time nature of electricity production and consumption (which occur within milliseconds of each other), and the difficulty of storing electricity. They note that policymakers can either set minimum-reserve margins or subsidize capacity with an up-front $/kW-year payment for capacity. In principle, the two approaches should yield the same outcome.

NERC (1998a) raises concerns that “few, if any, customers understand the implications of contracting for other than firm power supplies and firm transmission services.” Because of the long tradition of ample supplies and the use of interruptible rates to offer implicit discounts to large industrial customers, these customers are used to very few interruptions in service. Indeed, industrial customers, when interrupted, often are angry. Thus, it is an open question how customers will respond to real-time pricing. In addition, only a few electric utilities (e.g., Georgia Power) have much experience and a clear understanding of whether and how customers might respond to real-time pricing.

ISO APPROACHES

Bulk-power operations in California are split between the Power Exchange (PX) and the ISO. The PX runs day-ahead and hour-ahead (now, day-of) energy markets. In addition, the ISO operates a real-time energy market to balance generation and load during each hour. Neither the PX nor the ISO specifies installed-capacity requirements for market participants in California. And neither entity operates an installed-capacity market.

Between April 1, 1998, and March 31, 1999, the weighted average price of electricity in the PX day-ahead market was $26.6/MWh. For 12% of the hours, prices were at or below $10/MWh. At the other extreme, prices were at or above $100/MWh for 1.1% of the hours, with prices ranging as high as $200/MWh; these prices are well above the marginal costs of the most expensive units in California and reflect the pure capacity price shown in Fig. 4.

The California ISO (1998) points to the number and size of the proposed power plants in California (16 projects with a total capacity of more than 10,000 MW as of spring 1999) as evidence that competitive markets for capacity can work.

Nationwide, as of October 1998, developers had announced plans to build 109 merchant plants with a total generating capacity of over 56,000 MW (Thurston 1999). (According to the

*These societal benefits might include avoidance of the looting and violence that can erupt during a major blackout and the maintenance of electrical service to vital societal functions, such as hospitals, police and fire stations, traffic lights, and airport traffic-control systems.
Electric Power Supply Association, the total capacity of announced merchant plants had increased to almost 69,000 by February 1999.) Although many of these plants will likely not be built because of problems with siting, state approval, financing, or transmission access, these plans suggest that regulatory mandates are not needed to bring forth new generating capacity in the United States.

The PJM (1998a) *Reliability Assurance Agreement* (RAA) establishes the obligations of all LSEs within the PJM control area to provide the amount of installed generating capacity that PJM determines is needed to maintain reliability. The PJM Reliability Committee determines the forecast pool requirement, the reserve margin for the PJM Control Area required as part of this agreement. The RAA “is intended to ensure that adequate Capacity Resources will be planned and made available to provide reliable service to loads within the PJM Control Area, to assist other Parties during Emergencies, and to coordinate planning of Capacity Resources consistent with the Reliability Principles and Standards.”

The PJM Reliability Committee determines the forecast pool requirement for capacity resources using “probability methods” and establishes criteria for use of capacity resources during emergencies. The forecast pool requirement is intended to “ensure a sufficient amount of capacity to meet the forecast load plus reserves adequate to provide for the unavailability of Capacity Resources, load forecasting uncertainty, and planned and maintenance outages.” The focus is on the peak season, which for PJM overall is the summer.

In October 1998, PJM (1998b) established the monthly PJM Capacity Credit Markets to allow PJM market participants to buy and sell capacity credits to meet their obligations under the RAA. The markets for January through May 1999 cleared 5000 MW at prices that ranged from $50 to $80/MW-day. The average price during this 5-month period was $65/MW-day (equivalent to $24/kW-year, assuming that capacity is valued equally for every day of the year). PJM began daily markets (conducted a day ahead) in installed capacity in January 1999. For the four months of January through April 1999, the daily price averaged less than $5/MW-day, far below the prices in the monthly markets. (It is not known whether these price variations reflect seasonal differences, differences between daily and monthly markets, or lack of familiarity with these new products.)

The New York ISO (1998) proposal for installed-capacity requirements is similar to the PJM approach. New York explains clearly the purpose of its capacity requirement: “Adequate resource capability shall exist in New York State when, after due allowance for scheduled and forced outages and scheduled and forced deratings assistance of interconnection with neighboring Control Areas and regions, and capacity and/or load relief from available operating procedures, the probability of disconnecting firm load due to resource deficiency will be, on average, no more than once in ten years.”

The ISO New England (1997) approach differs from the PJM and New York approaches in that New England has two capacity components: monthly installed capability and hourly
The difference between installed and operable capability appears to be inoperable capability. Operable capability refers to “any generating unit or units in any hour … which is operating or available to respond within an appropriate period to the System Operator’s call to meet the Energy and/or Operating Reserve and/or AGC [automatic generation control] requirements of the NEPOOL [New England Power Pool] Control Area ….” New England market participants are required to bid all their operable capacity in excess of their obligations into the hourly operable capability market.

It is unclear why New England requires two capacity markets in addition to the energy and ancillary-services markets. Our discussions with several people in the region suggest that the two markets are historical artifacts and that, within a few years, one or both will be eliminated.

ISO New England (1998) noted that “NEPOOL has had an ICAP [installed-capability] requirement since its inception. This requirement has been important in maintaining reliability in New England for over 25 years.” These comments were in response to criticism from Cramton and Wilson (1998), who had been hired by ISO New England to review New England’s proposed market rules and were quite critical of these rules. Specifically, they wrote about the installed-capability requirement:

This holdover from an era of regulation is unique in the electricity industry, which is the only one that does not expect suppliers to cover fixed costs, such as capital and maintenance, from the market price of its output. … The capacity markets are a holdover from the regulated setting, when capacity decisions were not made in response to price expectations. In the transition to a competitive market, the capacity markets may serve a useful role in coordinating investments in capacity. However, once competitive electricity markets are established in New England, it would be appropriate for the capacity markets to terminate.

New England may maintain both installed- and operable-capability requirements because the installed-capability requirements are largely independent of availability. The installed-capability requirements relate primarily to “iron in the ground” without regard to the ability of that unit to operate any time soon. For example, the three large Millstone nuclear units were out of service for 18 months or longer, during which time they continued to qualify as installed capability in New England.

This discussion of the need for installed- and operable-capability markets raises the difficulty in determining what to include as installed capacity. Over what time period should generating-unit availability be measured? Should availability be determined on a daily, monthly, seasonal, or annual basis? Because the need for capacity is generally greatest during winter and summer peak periods, it may be most important to measure availability during those time periods.

The difference between installed and operable capability appears to be inoperable capability. Because no one should want to purchase inoperable capability, these two markets may be redundant. In addition, operable capacity seems to duplicate the real-power ancillary services.
QUANTITATIVE ANALYSIS OF GENERATION ADEQUACY

OAK RIDGE COMPETITIVE ELECTRICITY DISPATCH MODEL

We used ORCED to analyze and quantify these generation-adequacy issues (Hadley and Hirst 1998). ORCED is a simple strategic planning model that simulates the operations of, and resultant prices and producer profits from, competitive bulk-power systems.

The ORCED version used here consists of one region (within which transmission is unconstrained) with 50 generating units. Each generating unit is characterized by its heat rate, fuel type and cost, variable and fixed O&M costs, and capital costs. The model distinguishes between plants already online, for which capital costs are sunk, and plants that have not yet been built, for which capital costs are fully avoidable.

Consumer loads are represented by two load-duration curves for offpeak and onpeak seasons. Consumer responses to changes in electricity prices are represented by three input demand elasticities: (1) an overall elasticity that adjusts annual consumption up or down on the basis of decreases or increases in overall electricity price, (2) a time-of-use elasticity that changes the shape of the load-duration curve in response to changes in hourly spot prices, and (3) an unserved-energy elasticity that adjusts demand down during those periods when unconstrained demand would otherwise exceed capacity, used to calculate the market price at which supply and demand equilibrate. The current analysis uses only the third of these three elasticities.

ORCED is both an optimization model and a simulation model. We used ORCED to minimize the avoidable cost of electricity production for a given year subject to various constraints. (Additional analyses of required reserve margins vs reliance on markets should use dynamic models that analyze generation construction and dispatch over several years, rather than the single year considered here.) The constraints limited the megawatt capacity of new generation that could be constructed and ensured that no generating units lost money (i.e., every unit earned revenues greater than or equal to the unit’s avoidable costs). Given the set of generating units online, the simulation portion of the model then dispatched these units in least-cost order for the analysis year.
STRUCTURE OF ANALYSIS

We ran two sets of cases with ORCED for this project. In the first set, we fixed the unserved energy elasticity at 0.05 and ran several cases with different values of reserve margin. Because ORCED deals with energy and not with ancillary services, these reserve margins should be increased by at least 5 percentage points to reflect the need for generating capacity for regulation, spinning reserve, and supplemental reserve (Hirst and Kirby 1998). We set up these model runs to minimize the avoided cost of electricity production, taking into account both the construction of new generators and the operation of the existing fleet. We then added an annual capacity payment (in $/kW-year) to ensure that the most unprofitable unit just broke even. This capacity payment was determined by dividing the monetary losses for each generator (for those generators that lost money) by the availability-adjusted capacity of each generator. Given the required amount of installed capacity, the payment was then set equal to the highest dollar-per-kW loss to ensure that no generator lost money. This capacity payment (which ORCED caps at the carrying cost of a new combustion turbine, around $60/kW-year) is then added to the price of electricity that consumers pay.

In the second set of cases, we varied the unserved-energy price elasticity and let the model determine the “optimal” reserve margin. Here, too, we set up these model runs to minimize the avoided cost of electricity production, taking into account both the construction of new generators and the operation of the existing fleet. We added an O&M adder (expressed in $/kW-year) for those units that operate less than 10% of the hours. We added this factor (which ranged between 0 and about $2/kW-year) to ensure that plants operating for only a few hours a year would recover their avoidable fixed costs. (ORCED converts the adder to an energy-price premium paid to those units; the premium increases as capacity factor declines below 10%). The rationale for including this O&M adder is the same as that used to justify inclusion of the capacity adder in the fixed-reserve margin cases—to guarantee that no generator loses money.

---

1In a project for the U.S. Environmental Protection Agency, we are conducting additional ORCED analyses of generation adequacy. These analyses, under way at Oak Ridge National Laboratory, are examining the effects of higher natural gas prices, lower system load factors, and overall time-of-use demand elasticities on mandated reserve requirements vs reliance on market decisions concerning construction of new generating capacity.

6An unserved-energy elasticity of 0.05 (~0.05, to be precise) means that a 1% increase in the price of electricity cuts demand by 0.05%. Doubling electricity price cuts demand by 3.5%. This elasticity determines how high the spot price of electricity must rise to reduce demand so that demand equals online supply.

7Alternatively, we could have run the model to minimize the price of electricity or to maximize generator profitability. These alternative optimization criteria led to results that were, in our view, unreasonable. For example, minimizing the price of electricity led to the construction of many baseload units, which, in turn, led to major earnings losses for the owners of these units.

3This no-loss constraint is essential in competitive electricity markets. Were a unit to lose money continuously, it would go out of business, which would reduce the amount of installed capacity below the minimum specified. This problem is not solved by a change of ownership because it is a function only of operating, not capital, costs.
RESULTS

Our base case is an electric system with peak demand of 7000 MW and annual energy consumption of 38,600 GWh, yielding a system load factor of 63%. With an unserved-energy elasticity of 0.05 and a reserve margin of 5%, the shares of generating capacity and energy are: coal (39 and 50%), gas (33 and 25%), nuclear (13 and 17%), hydro (11 and 8%), and oil (4 and 0%). All the new generating units (960 MW) are gas-fired combined-cycle units. The annual average price of electricity is 2.89¢/kWh (ranging from 0.5 to 22.8¢/kWh throughout the year), and the avoidable cost of electricity is $773 million.

We ran cases with reserve margins ranging from 1 to 15% (Fig. 5). Model results show, as expected, that the total cost of electricity production increases at very low and at very high levels of reserve margins. For the assumed unserved-energy elasticity of 0.05, the cost curve has a broad minimum at around a 5% reserve margin. As the required reserve margin increases beyond the optimum value, the total price of electricity (i.e., the spot price of energy plus the capacity payment) increases. However, the spot price of energy declines and the capacity charge increases as the required reserve margin increases. Thus, setting the reserve margin too high drives a substantial wedge between energy prices and the total price that consumers see, undercutting the operation of the energy market. In addition, higher specified reserve margins lower peak period prices and the volatility of prices during the year. As the required reserve margin is increased from 1% to 5% to 10%, the maximum spot price declines from 28¢/kWh to 23¢/kWh to 19¢/kWh.

Electricity costs and prices are high at very low values of reserve margin because the amount and cost of unserved energy are high. At the other end of the spectrum, prices and costs are high because of all the “excess” capacity that must be supported through the capacity payment.

As the specified reserve margin is increased, the amount of new generating capacity brought online also increases. For these cases, all the new capacity is gas fired. Because this new capacity is very efficient, it accounts for larger shares of energy production than of generating capacity. For example, at the “optimal” 5% reserve margin, new gas-fired generation accounts for 14% of generating capacity and 19% of energy production. For reserve margins of 10% or below, all the new capacity is combined-cycle units; at higher values of reserve margins, some of the new capacity is combustion turbines. At high values, the capacity factor for generation declines. For low-capacity-factor operation, combustion turbines are a more economical choice than combined-cycle units because of their low capital costs (offset by their higher operating costs).
We then ran ORCED with no required reserve margin but with different values for the consumer response to price changes when unconstrained demand would otherwise exceed supply (the third elasticity factor discussed above). The results show substantial benefits from encouraging at least a minimal level of consumer response to high prices (Fig. 6). An increase in elasticity from 0.01 to 0.04 cuts costs and prices by 7% and 15%, respectively. The results also show that, beyond an elasticity of about 0.04, there is little additional benefit to greater consumer response to price changes. At higher levels of elasticity, costs and prices decline very slowly. As demand elasticity increases, the maximum spot price and price volatility decrease. As the elasticity increases from 0.02 to 0.05 to 0.10, the maximum spot price declines from 339¢/kWh to 38¢/kWh to 13¢/kWh (Fig. 7).

Figure 7 shows the reserve margins chosen by ORCED as a function of demand elasticity. As expected, this market-determined reserve margin decreases as elasticity increases. The reserve margin is about 5% at an elasticity of 0.05, consistent with the results shown in Fig. 5. Figure 7 also shows that the price of unserved energy, the market price of electricity at times when supplies are not sufficient to meet unconstrained demand, declines steeply as the elasticity increases. However, at very low values of elasticity, the price can be quite high; at an elasticity of 0.015, the price of electricity would have to be 629¢/kWh to ensure that demand declines enough to match supply.
Fig. 6. ORCED results showing electricity prices and costs as functions of the unserved-energy price elasticity.

Fig. 7. ORCED results showing market-determined reserve margin and unserved-energy (peak) price as functions of the unserved-energy elasticity.
DISCUSSION

What can we conclude concerning the two primary approaches to ensuring that enough generating capacity is available so that customers will not be involuntarily disconnected from the grid? Those most concerned about reliability note that requiring minimum planning reserves (1) ensures that “enough” generation will exist; (2) uses an approach that worked well in the past; (3) reduces the volatility of electricity prices; (4) protects customers, most of whom do not want to deal with the complexities of time-varying prices, from such volatility; and (5) ensures that the positive externalities associated with extra generating capacity are maintained. Those most concerned about economic efficiency and development of competitive markets for electricity counter that (1) the amount of generating capacity that is “enough” depends on customer response to prices, which varies across customers and customer classes; (2) because an approach worked well in a vertically integrated, monopoly-franchise industry is no assurance that it will work well in a deintegrated and competitive industry; (3) price volatility sends important economic signals to consumers and producers concerning when and how much electricity to consume and produce; (4) only a small fraction of loads needs to be price-sensitive to equilibrate demand and supply and to eliminate the need for mandated planning reserves; and (5) no public benefits are associated with generation adequacy beyond the private benefits.

In addition, the market proponents favor market decisions on generation investment because it (1) lets markets make both capital and operating decisions, which is why we are creating competitive electricity markets in the first place; (2) encourages customer participation in the provision of reliability services; and (3) will yield lower average electricity prices and costs to consumers. The reliability proponents respond that: (1) reliability is in part a public good that cannot be left entirely to the self-interests of market participants; (2) there is too much uncertainty about whether, when, and by how much customers will reduce demand in response to high spot prices to consider demand-side responses a resource for reliability; and (3) the costs of major outages are so high that they wipe out any savings associated with lower electricity prices.

The ORCED analyses suggest that centralized decisions concerning the amount of generating capacity to maintain may often be wrong, yielding either too much capacity or not enough. So many factors affect the “optimal” amount of generating capacity (e.g., the prices of fossil fuels, the amount and types of generating capacity already online, and customer load shapes and price elasticities) that it is very difficult for central planners to make the right choice. However, the ORCED results (Fig. 5) show a broad optimum, ranging in this case over several percentage points around the true optimum.

On the other hand, relying on the actions of consumers and suppliers in response to time-varying spot prices works well only if consumers can and do respond to high prices. Therefore, electricity policymakers should encourage demand-side experiments and investments to ensure that, when prices rise, customers will be able to respond.
CONCEPTS

The basic function of transmission is to interconnect loads and generators. Originally, this function was accomplished within a vertically integrated utility structure. Transmission interconnections between utilities were constructed to increase supply options and reliability. Utilities discovered that the same interconnections designed to enhance reliability by enabling the sharing of reserves were also useful for facilitating economic transactions (based on differentials in the cost of energy production at different generators). Of course, even the use of transmission to enhance reliability was driven by economics; it was cheaper to share reserves than to build redundant generation-reserve capability.

When there is insufficient generation available to maintain the required generation/load balance, load will be reduced to match generation, generation will be increased to meet load, or the system will collapse. (Alternatively, if the generation inadequacy is within only one control area of an interconnected system, the deficient control area can temporarily lean on the interconnection to restore its balance.) The concept of adequacy is much more complicated for transmission.

The California ISO (Miller 1998) lists six reasons that transmission enhancements may be required:

- Interconnect generation or load (e.g., build a radial line from a new generator or load to the transmission system)
- Protect or enhance system reliability (e.g., replace older, less reliable equipment with newer, more reliable equipment)
- Improve system efficiency (e.g., replace high-loss equipment with lower-loss equipment)
- Enhance operating flexibility (e.g., add switching capability)
- Reduce or eliminate congestion (e.g., add new transmission lines or increase the capacity of existing lines)
- Minimize the need for must-run contracts (e.g., add transmission lines or reactive support at locations that depend on a single generator)

It is difficult to separate reliability from commerce in determining the motivation for a particular transmission project. Many real-world enhancements address multiple needs. An additional line bridging a congested interface would probably reduce congestion, increase
reliability, and improve efficiency. It might also increase operating flexibility and minimize the need for must-run contracts.

Transmission can be so inadequate that it is not possible to meet the load requirements in a given location, which would be a reliability concern. Load must then be curtailed to prevent a collapse of the power system. More commonly, when transmission is inadequate, it means that the generation dispatch must be constrained (i.e., economically suboptimal) to maintain system security. The result is a higher price on one side of the constraint and a lower price on the other. In the extreme, specific units must run to maintain security. These “reliability-must-run” units might have substantial market power. Consequently, such units often run at the system operator’s discretion and receive a regulated payment (Jurewitz and Walther 1997).

Transmission adequacy has two parts:

- There must be sufficient transmission to support balancing load and generation given known and expected outages.
- It is desirable to have enough transmission so that competitive generation markets can function.

Historically, generation and transmission were planned on an integrated basis by a vertically integrated utility under state regulatory oversight. The traditional process met the first objective (balance generation to load) but had little need to address the second objective (facilitate competitive generation markets). A vertically integrated utility strives to have sufficient transmission to economically supply its load. Power system planners forecast load patterns and generation-resource availability. Historical performance, including the load at each bus, was used to create detailed models of the electric system for peak and off-peak conditions. Models, data, analyses, and transmission plans were coordinated with neighboring systems.

A basic question should be addressed at the outset. Because transmission costs only one tenth as much as generation, why not build enough transmission so that it never constrains generation markets? This approach is appealing to engineers who like systems with flexibility and sufficient excess capacity to meet unexpected needs and to market participants who want to buy and sell power over large geographic regions. However, the construction of new transmission lines is often opposed by local residents and landowners and is therefore politically difficult to achieve. And regulatory rules may not permit utilities to recover fully the costs of such “overbuilt” systems. Who should pay these costs is also unclear. Finally, transmission systems that appear to be robust and flexible are designed to service a specific set of expected generation and load patterns. Attempting to facilitate all possible combinations of generation and load would overwhelm any realistic transmission system.
Flows within a transmission system depend on the configuration of the system, the generation injections, and the load withdrawals. There is little ability to control flow within the network other than by removing elements (taking transmission lines out of service) or by changing the generation injections. Planners test system adequacy by modeling performance (line flows and bus voltages) under a full range of expected load, generation, and contingency conditions. Limits on the acceptable generation dispatch range are determined for each set of operating conditions. The planners then make a judgement as to the adequacy of the transmission system.

Sometimes the judgement is straightforward. When the load being served by a radial line exceeds the line’s capability, the transmission system must be enhanced, or local generation must be added. Similarly, when a point in the network cannot be served without exceeding the remaining lines’ emergency ratings during a contingency, the system must be enhanced. In both cases, the need is characterized as a reliability requirement. One or more enhancements are designed. The need is explained to the regulators, and an enhancement is selected. The project is built, its costs are included in the transmission rate base, and these costs are borne by all customers. Eminent domain is generally available to electric utilities to facilitate the acquisition of any needed right-of-way.

The need for enhancement is less clear in the more common case when inadequate transmission results in constraints on economic dispatch of generators rather than forcing curtailment of load. The complexity is twofold. First, the increased cost that will result from constraining the dispatch is generally an operating cost whose magnitude depends on the number of hours a year the constraint exists and the relative costs of the generators involved. At the same time, the transmission-enhancement cost is primarily a capital cost, which must be recovered over several decades. Which solution is the more economical depends on a number of factors, such as fuel costs, the cost of capital, and the expected locations and magnitudes of load growth and generation construction. Second, the need is less clearly reliability based. There may be opposition to using eminent domain to address an economic problem, especially if much of the economic benefit accrues to electricity consumers or providers in another state. To illustrate, the Wisconsin Public Service Commission (1998) wrote: “Consideration of transmission constraints in northern Illinois points to an important lesson: significant improvement in transmission transfer capability into eastern Wisconsin will require the construction of new facilities in other states. Ultimately, transmission system limitations on the scale considered in this study must be regarded as a regional problem and regional solutions must be fostered.”

---

1 Phase-angle regulators and flexible AC transmission systems (FACTS devices) provide limited, and expensive, control of flows at a few specific locations.

2 Controlling the precontingency (normal conditions) injection of power by generators can ensure that the post-contingency line flows are within emergency ratings of all lines.
Restructuring further complicates this situation. Organizationally separating transmission from generation means that transmission planners must forecast the commercial actions of generation owners (i.e., the timing, location, and size of new generating units) as well as load growth. The generation forecast must account for the operation of generators under market conditions rather than simply the cost differences that were considered when utility planners optimized over generation and transmission. The effects of these decisions are different as well. A generator that is constrained out of the market for too many hours a year may be driven out of business because it cannot meet its fixed costs. Alternatively, a generator that must run to meet local voltage or other reliability needs may have unacceptable market power and be denied market-based rates. If market conditions change such that a transmission project is no longer economical, on the other hand, the cost is borne by the customers of this regulated asset.

LESSONS FROM THE GAS INDUSTRY

Restructuring started earlier in the gas industry so there has been more time to develop innovative solutions. As it turns out, the simpler nature of gas transportation makes the problem easier and diminishes the relevance to electricity (Coxe 1999). It also makes it easier to see fundamental relationships that are important to both industries.

Consider the proposed pipeline between Texas and Chicago shown in Fig. 8 (Barcella 1999). Even if there are no other pipelines between Texas and Chicago, the new pipeline is not a monopoly if both Texas and Chicago have other choices. Texas can sell gas to New Mexico, and Chicago can buy gas from Canada, so the new pipeline does not have to be regulated as a monopoly.

How would the gas market value this pipeline? The market value is not related to the cost of the pipeline. It depends on the relative price of gas at the two ends of the pipe. If the current market price of gas is $8 in Chicago and $5 in Texas the value of transporting gas is $3, today. When considering a pipeline of significant capacity, it is also necessary to estimate how the pipeline will affect prices at both ends. If the pipeline will buy enough gas in Texas to

![Proposed pipeline links locations with different prices.](image)

"Vertically integrated utilities operated a portfolio of generation resources and were concerned about its overall profitability. In competitive generation markets, each unit must stand on its own as a profit-making enterprise."
raise the price from $5 to $6, and if it will sell enough gas in Chicago to drop the price from $8 to $7, the value of the pipeline will drop from $3 to $1.

Is there a need for central planning of the pipeline network? No. Publicly available location-based market prices provide sufficient information for investors to determine what projects are needed, when, and where. Investors are free to forecast prices and perform engineering analysis to determine costs. Projects that show sufficient promise of turning a profit will likely be built. Regulators need only ensure that projects are sufficiently reasonable that they will not become a public burden and worry about land-use issues (no small problem in itself).

Interestingly, market forces can still facilitate pipeline expansion if pipeline companies themselves remain regulated and restricted to a cost-based rate of return. The pipeline company itself may be unlikely to invest in new capacity. A cost-based rate of return that caps the pipeline’s transportation price will limit the profits that can be made during peak use when prices in Chicago greatly exceed prices in Texas. Regulations will also discourage customers from buying pipeline services during off-peak times when the difference between the Chicago and Texas prices is less than the pipeline’s capped rate. Other investors, however, could sign long-term firm contracts with the pipeline, guaranteeing the pipeline an income during low-flow times but keeping the profits during times of high flow and high price differences. The investor could be one or more loads or producers or third parties with no other market interest.

Differences between natural gas and electricity limit the applicability of the rules developed for one market to the other. First, while flow through an individual natural gas pipeline can be controlled independently of the rest of the network, it is not yet economically practical to control the flow on individual transmission segments. A transmission line cannot be used as an active market participant transporting a controlled amount of power from one market location to another. Conversely, since the flows throughout the network are interdependent, a party with economic interest in the power markets cannot be allowed to manipulate portions of the grid to create congestion by denying access to existing transmission capacity. Second, the need to instantaneously balance the production and consumption of electricity necessitates constraining operations of the network based on what might happen during and after the next contingency rather than on current conditions. Therefore, the value of a specific transmission element is not necessarily reflected in the current flow. The ability to store gas, in tanks and in the pipelines themselves, greatly alleviates this problem for gas. Third, economies of scale are very large with electric transmission projects. It is much cheaper to install all the capacity a transmission line is likely to need initially than it is to retrofit the line later. With gas, on the other hand, it is not uncommon to initially operate a line at one pressure and later raise the pressure to obtain additional capacity. Consequently, a new electric transmission line is much more likely to completely eliminate congestion and the locational difference in price between the two ends of the line, eliminating any congestion revenue it might otherwise collect.
Still, there are important lessons that can be learned from the gas analogy (Table 1). The market values transportation based on the difference in real-time prices at the line ends. Just as the prices vary in time, so does the value of transportation. Price differences help direct current operations and provide signals concerning where transmission investment is needed. Investment results because of an expectation to profit, either from extracting a toll for moving a product from a region with one price to a region with a higher price (transporter profit) or by modifying the regional price for the product (generator or load seeking to raise or lower the price they face).

Table 1. Comparison of gas-pipeline and electric-transmission-line characteristics related to congestion and capacity expansion

<table>
<thead>
<tr>
<th></th>
<th>Gas pipeline</th>
<th>Electric transmission line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market valuation</td>
<td>Commodity price differential at ends</td>
<td></td>
</tr>
<tr>
<td>Control of line flows</td>
<td>Line flows can be controlled independently</td>
<td>Individual line flows depend on generation and load patterns and all other line impedances; loop flows and parallel paths are problems</td>
</tr>
<tr>
<td>Withholding capacity from the market</td>
<td>Acceptable</td>
<td>Cannot be allowed because of the impact on other line flows</td>
</tr>
<tr>
<td>Incremental capacity additions to match market requirements</td>
<td>Significant ability to add capacity incrementally</td>
<td>Generally cheaper to initially install full capacity (i.e., overbuild)</td>
</tr>
<tr>
<td>New capacity impact on price differentials at terminals</td>
<td>Manageable</td>
<td>Often eliminates price differential</td>
</tr>
<tr>
<td>Investment recovery through market pricing of capacity</td>
<td>Likely</td>
<td>Unlikely</td>
</tr>
<tr>
<td>Capacity planning</td>
<td>Individual</td>
<td>Centralized</td>
</tr>
</tbody>
</table>

LOCATIONAL PRICING AND CONGESTION CONTRACTS

Competitive markets for generation and retail services can be accommodated with monopoly management of transmission operations and investment. The monopoly could take on the obligation to provide unlimited transmission service for everyone. The monopoly would
make investments and/or pay for redispatch to manage congestion. The National Grid Company in England and Wales works essentially this way. All the usual problems with regulating and providing appropriate incentives to a powerful monopoly still exist.

There is an alternative to complete monopoly management of transmission expansion. In spite of the difficulties and complexities, several schemes have been proposed for pricing transmission based on locational electricity prices (Hogan 1998; Oren 1997; Tabors and Galindo 1999). The inability to control individual line flows makes completely independent transmission projects and markets impossible. The ability to create congestion and drive locational prices apart through inappropriate operation of the transmission system creates a need to separate physical control of the transmission system from the rights to profit from congestion. Thus, the network needs a system operator, independent of market outcomes, to manage real-time operations. Flows will distribute themselves among the transmission lines based on the locations and magnitudes of generation and load as well as the impedances of the individual transmission elements. Therefore, it would not be feasible for a transmission owner to place a “toll booth” on a particular line and levy a charge for every transaction that used that line. But something that is economically similar can be developed.

Hogan (1998) describes transmission-congestion contracts (TCCs) that are equivalent to perfectly tradeable physical transmission rights: “With such contracts to allocate transmission benefits, it would be possible to rely more on market forces, partly if not completely, to drive transmission expansion.” In Hogan’s scheme, transmission service is priced partially on the nodal price differences. Nodal price differences result when transmission is congested and there is insufficient transmission capacity to move lower-cost power to higher-cost locations.

Congestion is not related to the actual flows on lines. Congestion occurs when security-constrained dispatch requires modification of the economic dispatch. This situation occurs most frequently as the result of contingency analysis rather than because of steady-state line flows. The generation dispatch is modified because a line will overload if a specific contingency occurs (e.g., a generator or transmission line trips) (Boucher et al. 1998). Because there is no time to take corrective action to prevent cascading failures, it is necessary to preemptively modify the generation dispatch. It is this off-economic dispatch that results in locational price differences. (Losses also cause locational price differences but have a much smaller impact and are easier to deal with than congestion and are ignored here.)

---

*Oren proposes active rights and link-based contracts instead of TCCs. Tabors argues that most of the effects of congestion can be accounted for on a zonal, rather than nodal, basis, which facilitates ex ante pricing and ex post settlement. In either case, the detailed distinctions may be important to implementation but do not affect the conclusions of this discussion. The Hogan, Oren, and Tabors schemes all encompass the same basic requirements.*

29
Why is congestion so much more important a problem now than it was a few years ago? The traditional vertically integrated utilities accounted for transmission constraints when they made their daily operating (unit-commitment) plans. Thus, they used their generating resources in a way that would not overload the network. In today’s increasingly competitive environment, suppliers schedule resources without a detailed knowledge of transmission constraints.

TCCs do not give the contract holder the right to move an amount of power over a particular line or between a particular pair of locations. Instead, they give the holder an income based on the price difference between the two locations. This achieves essentially the same result without disrupting the actual generation dispatch. To see how this works, consider the generator G and load L in Fig. 9. L contracts for 200 MW from G at $40/MWh. How will the generator get the power to the load?

G sells its output through the PX at the locational spot price. L purchases power through the PX at its locational spot price. The TCC operates to turn these two spot transactions into a firm power contract with guaranteed transmission. If the price at both L and G rises to $50/MWh, L pays $50/MWh, G receives $50/MWh, and G pays L $10/MWh. The net result is the same as L paying G $40/MWh. If congestion between G and L results in the price at L rising to $60/MWh, G’s TCC results in a $10/MWh payment ($60 – $50/MWh) to G. L pays the PX $60/MWh, G receives $50/MWh from the PX. G then pays L the $10/MWh TCC payment and the $10/MWh excess profit it received from the PX ($50 – $40/MWh). Again, the net result is a transaction between G and L at $40/MWh. The PX retains none of the TCC revenues.

**Fig. 9.** TCCs are the economic equivalent of firm transmission contracts between locations with different power prices.
TCCs could be awarded in various ways. Existing capacity could be divided to reflect historical rights. Alternatively, TCCs could be auctioned off. While the allocation process may be very important to the individuals, only the fact that the TCCs are allocated is important to the system. Once allocated, TCCs allow markets to appropriately value and allocate scarce transmission among competing users.

In defining the TCCs, making sure they are simultaneously feasible is essential. This is done by modeling the transmission system. In all cases, the ISO or a neutral analysis party determines the transfer capability from each bus to every other bus. This analytically determined transfer capacity is then auctioned or allocated, no matter where the constraint actually lies. The constraint’s impact on the transfer capacity between each pair of buses is what is important.

Potential purchasers can make price vs quantity bids of what they are willing to pay to acquire a TCC between each pair of buses. To select the best combination of bids, the ISO should optimize consumer welfare. An owner of a TCC now has a specific MW of capacity between two buses. Whenever the spot price deviates at the two buses, the TCC owner gets the differential price times the TCC congestion capacity. Note that the TCC owner may have no idea who is transacting to create the congestion. The congestion may result from commercial flows between totally different buses. Only the system operator has sufficient information to administer the TCC payments. To Hogan, TCCs are even better than long-term physical contracts because there is also a secondary market, the pool, to accommodate extra load or extra generation.

Looking only at the locational marginal prices and the defined capacities between specific locations makes it possible for system users to ignore the physical details (the system operator must remain keenly aware of the physical system). In fact, a market can be administered that converts the physical network into a hub-and-spoke system, a much simpler trading structure (Hogan 1998).

The system operator should not be allowed to keep the excess TCC payments (in general, there will be more congestion payments collected from locational price differentials than are paid out to congestion-contract owners). If congestion payments remained with the system operator, this would give it an incentive to create congestion. Instead, excess congestion payments should be used to offset the fixed costs of transmission.

TCCs AND TRANSMISSION INVESTMENT

TCCs can help allocate scarce transmission among competitive users in the short term and they can help with transmission investment in the long term. Congestion contracts covering the additional capacity created by transmission enhancements are awarded to the investor. The system operator uses the same system models that allocated the original TCCs to determine
what additional TCC capacity is created by the transmission enhancement. All existing TCC rights are preserved, and the investor gets the new TCC rights.

An interesting problem arises however. Note that the investor gets the new MW capacity but congestion prices are likely reduced, possibly to zero. So, although everyone’s rights are preserved, the value of those rights can be wiped out. Hence, congestion contracts will most likely be tied to the generators or loads that profit from reduced congestion rather than to investors speculating in congestion or building transmission for profit. Although generators or loads lose the value of their congestion contracts, they gain as much or more from reduced power prices at the receiving bus or increased prices at the sending bus.

This impact of significant transmission enhancement on locational prices creates a free-rider problem that may make it difficult to get investors to invest. Each party is better off if someone else makes the investment. The problem is compounded by the scope and scale problem. It would be desirable to invest only enough to cover the exact amount of transmission capacity required by the individual needs. Then anyone else will cause congestion and there will be a price differential that the investor’s TCC will collect rent from. But that is hard to do. It is generally much cheaper to oversize a line when constructing it than it is to come back later and upgrade it. Similarly, in an interconnected system, economies of scope often mean that an enhancement removes constraints on a number of transmission paths. Customers using one path may choose to wait until the customers using another path make the investment. The longevity of transmission equipment (roughly 50 years or more) adds a further complication.

Even if all of the current beneficiaries could agree on the need for a transmission enhancement, they are unlikely to follow through. A user of the transmission facility would have to be able to commit for decades of benefits to pay for the investment. Since the enhancement immediately becomes a sunk cost, providing the benefits at nearly zero incremental cost, there is no ability for the current user to sell its share in the investment if it no longer needs it unless the system becomes congested again.

In cases where it is not practical for specific users to invest to relieve congestion, it may be appropriate to rely on congestion pricing to identify economically sound investment. Once identified and approved by regulators, the transmission enhancement could be built, either by the existing transmission owner or by a third party. The enhancement could then be added to the rate base for all users of the system or for users of a portion of the system.

Eminent domain could become a problem, making transmission right-of-way harder to obtain for commercial reasons than for reliability reasons. A regulatory approval process could
address this problem as well. There has been little concern raised in the literature, however, so this may not be a real problem.

TRANSMISSION-PRICING PRINCIPLES

While congestion contracts may seem complex, several fundamental principles concerning pricing transmission in a restructured industry are generally accepted:

- No one can be allowed to withhold transmission rights; unused capacity is available for others to use.
- The system operator must be involved in the analysis of rights and the calculation of payments for transmission congestion and losses.
- The system operator must not be allowed to profit from congestion.
- Transmission prices should reflect differences in locational power prices.
- Transmission prices can send appropriate signals for the locations of new transmission and generation facilities as well as loads.
- Transmission prices can illuminate the need for transmission enhancement but they are unlikely to provide sufficient incentive to motivate investment without additional compensation.
- The remaining transmission costs will have to be allocated among users.

These principles, unfortunately, do not lead to a completely market-based expansion policy, as is possible with gas and most other industries. This situation leaves most enhancement decisions in the discouraging position of having many interested parties and no uniquely correct solution. To make matters worse, the optimization is often quite flat from an overall system perspective but can have dramatic impacts on individuals. That is, there are often multiple solutions that are, from the system’s point of view, equally good (i.e., low in cost). But the differences can be critically important to individual market participants. With a vertically integrated utility, differences in the least-cost solutions did not matter. The customers only paid the aggregated cost. Now it is very important because individuals prosper or starve based on the final decision. The selection procedure can be, or appear to be, arbitrary. For example, a generator located within a congested portion of the grid might be driven out of business if the congestion is relieved, a dramatic result for the owner of that unit. The amortized capital cost of relieving the congestion, on the other hand, may be only slightly lower than the off-economic dispatch cost, a small impact on the overall system. The fact that the decision to invest will, of necessity, be based on forecasts makes the problem worse. Finally, the analysis and decision will likely be made by individuals (the ISO and the regulator) that bear no market risk themselves.

*The ERCOT ISO (1998) discussed the need for the transmission provider to obtain a certificate of convenience and necessity from the Texas PUC in the shortest reasonable time and that an independent body, such as an ISO, should help expedite the approval process by assuring the PUC that the project is needed for regional reliability.*
“Appropriate incentives” are the key to make either Transcos or ISOs economically efficient, encouraging them to appropriately balance reliability and commerce.

Pierce (1999) notes that FERC’s transmission-pricing approach has three flaws: (1) it is based on average embedded costs instead of marginal costs; (2) it is based on the contract-path fiction instead of the laws of physics; and (3) it assumes implicitly that separate prices for each segment of the transmission grid can aggregate to efficient prices for regional transmission service. “FERC needs to abandon completely each of the transmission pricing principles announced in Order 888. … Conversely, a transmission pricing regime based on marginal cost will send price signals that accurately reflect relative scarcity, thereby encouraging an efficient pattern of investments in transmission and generating assets. Thus, for instance, if a transmission capacity constraint creates a significant price difference between one location and another, prospective investors will know that they must either invest in new generating capacity in a location that is on the right side of the constraint or invest in the new transmission capacity required to reduce the effects of the constraint.”

With transmission remaining largely regulated and with transmission-enhancement projects often eliminating locational price differences caused by congestion, the problem of allocating the cost of transmission enhancements to customers remains. Some argue that if the cost of the existing transmission system, which connects existing generators to loads, is paid for by retail customers, then it would be anticompetitive to require new generators to bear the cost of the transmission that brings their product to market, although it would be reasonable to require a new generator to pay for any special interconnection and transmission-enhancement requirements (Texas PUC 1998). Not surprisingly, potential owners of new generation argue that any additional cost could be included in the transmission rate base if the PUC agrees that the project offers unusual benefits to the system. Texas Utilities proposes a modified scheme in which new generators pay the total cost for facilities that accommodate them and do not increase transfer capability from one region to another. New generators would pay only a pro rata share of the cost of facilities that increase interregional transfer capacity (Texas PUC 1998).

TRANSMISSION-EXPANSION APPROACHES

As regional transmission entities (Exhibit 3) develop in different ways throughout the country, they are adopting and applying different transmission planning and investment decisions. Entergy (1999) and FirstEnergy (1999) recently applied to FERC to establish Transcos; neither filing says much about transmission planning and expansion. Both note the benefits of combining transmission ownership and operation in one entity. Entergy claims that:

… a Transco will be driven, through appropriate incentives* to minimize costs, maximize throughput, achieve efficient levels of congestion and reliability, and expand the transmission system when economically justified. Unlike an ISO, this alternative structure will retain the efficiencies gained by integrating the

*“Appropriate incentives” are the key to make either Transcos or ISOs economically efficient, encouraging them to appropriately balance reliability and commerce.
Exhibit 3. Regional Transmission Entities

During the past several years, various groups have proposed different kinds of regional organizations to operate today’s transmission systems and to plan and build expansions to those systems. These entities include regional transmission groups, ISOs, independent system administrators, transmission-owning companies (Transcos), and regional transmission organizations [RTOs, FERC’s (1998c and 1999) latest term for such organizations].

As of spring 1999, five ISOs operate in the United States (California, ERCOT, PJM, New York, and New England), and soon the Midwest ISO will join them. Several companies plan to create Transcos (including NSP and Alliant, a group of utilities called the Alliance Regional Transmission Organization, FirstEnergy, and Entergy). The primary difference between an ISO and a Transco is that an ISO directs the operation of transmission assets that it does not own, whereas a Transco combines ownership and operation of the grid assets in one entity. An additional distinction that some Transco proponents make is that a Transco will be a for-profit enterprise, whereas the five operating ISOs are all nonprofit entities.

The Alliance Regional Transmission Organization (1998a) favors the Transco structure because: “The combination of ownership and operation may be very appealing to some investors and experts in managing transmission who view the future transmission business as an independent business that creates value for customers and shareholders alike. Aggregating management and technical expertise to this singular and clear focus offers the best opportunity for coupling performance to value; it is not clear that this opportunity will exist in other structures where ownership and operation are separated.” Transco advocates suggest that the owner-operator status of the Transco will lead to a more efficient, integrated approach to transmission planning and investment than one that must be coordinated across multiple organizations (e.g., the ISO and all the transmission owners).

The Large Public Power Council (1998), on the other hand, argues that a not-for-profit Transco would better protect and balance the interests of all stakeholders, promote the public interest, have an open governance structure, lend itself to light-handed FERC regulation, and make it easier for municipal utilities to convey or transfer operations of their transmission to the Transco than would be the case with a for-profit Transco. The Council fears that a for-profit Transco could use its control over transmission to manipulate generation markets and increase transmission profits. The Transco could have an incentive to either “gold plate” or “tin plate” the transmission system, depending on how FERC set its rate structure. Any efficiencies gained by the not-for-profit Transco, on the other hand, would go to ratepayers, so less regulation would be required. The Council also notes that various restrictions prevent public-power entities from granting control of their facilities to for-profit organizations. The benefits of a large regional organization controlling the transmission system could be achieved, and include public-power facilities, by having a not-for-profit Transco.

The ISO-vs-Transco debate is not relevant to transmission adequacy because the transmission planning and pricing complexities discussed above apply to all regional transmission organizations.
operation of the system with the maintenance, engineering, construction, and restoration of that same system. Having the asset management portion of the transmission business working in tandem with, answering to the same management team, and driven by the same incentives as the operation portion of the business, will ensure that the system is operated, maintained, and expanded in the most efficient manner.

Western Interconnection

The Western Interconnection Coordination Forum (WICF 1998) was formed in 1996 to coordinate transmission planning throughout the Western Interconnection. WICF members include the three western regional transmission associations, the Western Systems Coordinating Council, the California ISO, the Committee on Regional Electric Power Cooperation (the PUCs in the region), and the Colorado Coordinated Planning Group. WICF requires and facilitates a common planning process throughout the West. The lengthy and involved WICF process for identifying and proposing projects is a very open one that allows anyone to propose a new transmission project. WICF assures that a thorough, traditional centralized planning process occurs. Planning data are collected from all members, and all members are invited to participate in the planning process for all proposed enhancements. A biennial transmission plan is produced, the first of which was published in May 1998, that attempts to address all of the transmission needs of all members. An outreach effort is conducted to obtain input from nonmembers. The final product is a coordinated transmission plan that includes input from a broad spectrum of stakeholders.

Although it is likely that identified projects will be pursued, neither implementation nor cost allocation were discussed in the biennial plan. The plan also did not consider alternatives to new transmission. The plan notes that the study results showing potential problems “does not mean that new transmission should be built to correct the identified problems. In general, reliability problems can be corrected by building additional facilities, by upgrading existing facilities, by adding or improving remedial action schemes, or by limiting path schedules on affected transmission paths.” Additional options include generation redispatch, new generation, and demand-side measures.

Midwest ISO

The Midwest ISO (1998) is under development (Volpe 1999). It will take the lead in developing a regional transmission plan that meets the need of all stakeholders, based in part on data and analyses provided by the member transmission owners. The transmission owners are then obligated to make a good-faith effort to design, certify, and build the facilities in their service areas that are part of the transmission plan approved by the Midwest ISO Board.

Customers that cause upgrade requirements must, during the ISO’s 6-year transition period, pay the annual carrying cost of the upgrades. Thus, during the transition, the customers
whose loads are located in a particular zone pay for those zone-specific upgrades. (After the transition period ends, such costs will be rolled into ISO systemwide rates.) Alternatives other than construction are analyzed for economic justification.

Alliance Regional Transmission Organization

The Alliance Regional Transmission Organization (1998b) is a work in progress. A recent proposal calls for the ISO to perform traditional central transmission planning. The ISO will identify transmission needs related to new loads, new generation, reliability, transmission-service requests, improvements to enhance efficiency, and congestion relief. (We doubt that transmission projects can be so neatly distinguished from each other in purpose.) The ISO must consult with all members and other interested parties. The ISO has the authority to order members to finance and build new facilities and to maintain existing facilities. If a project does not provide the desired results, the ISO will recover costs through scheduling charges to reimburse the transmission owner.

The ISO has the responsibility to serve any new load with transmission and to connect any new generator to the system. The ISO can compel transmission owners to build required facilities. The ISO is obligated to assist in obtaining regulatory approvals. The details concerning third-party construction and ownership of new facilities are unclear, but the existing transmission owners have, at a minimum, the right of first refusal on all new projects. Generators are responsible for any projects and their costs that are completely radial to their facility.

ISO New England

The interactions among NEPOOL, ISO New England, independent power producers, and FERC show how the region is addressing transmission expansion and also what FERC allows in assigning the costs of new transmission facilities to different entities.

The New England transmission system is largely uncongested internally but regularly congested on the interfaces to New York and Canada. The system operators relieve congestion by redispatching generation, with the redispatch costs shared among all loads within the region. New England has a postage-stamp transmission tariff with no locational signals, which provides no information to generators on where to locate new facilities. Historically, generators paid for interconnection facilities, and loads paid for the remainder of the transmission system. As in most regions, generation and transmission were jointly planned through an integrated, regional process dominated by the large investor-owned utilities. Now, with 30,000 MW of new generation being proposed (the existing system has 25,000 MW of generation), new generation is driving transmission expansion more than load growth (Tierney 1998).

---

1It is not clear whether or how an RTO could compel transmission owners, or anyone else for that matter, to build something. Presumably, FERC could compel such construction by a jurisdictional utility.
NEPOOL proposed a process that starts with system-impact studies conducted for each new generator. The ground rules for these studies require that the new generator be “fully integrated” with load, which means that the new generator must be capable of serving load anywhere in New England without reducing the ability of any existing generator to serve load anywhere in New England. Transmission, therefore, would have to be expanded to accommodate each new generator. The new generator would pay 50% of the cost, and loads would pay the remaining 50%. If the expansion, based on its $/MW of new generation, cost more than the system average embedded cost, the new generator would also pay the full above-average cost.

Further, NEPOOL was concerned that it has no authority to differentiate among market participants. “NEPOOL should not be in the business of speculating over market winners and losers” (FERC 1998a). So the proposed evaluation process treats new generators on a first-come, first-served basis and assumes that all existing generators and all other generators in the queue ahead of the generator being evaluated will be built and will operate. Further, redispatch was not allowed in the system-impact studies, only transmission expansion.

NEPOOL’s process might be reasonable for dealing with small changes in the generation market (with the possible exception of allowing redispatch for existing generators but not for new generators). It is clearly unreasonable for dealing with proposals that would more than double existing generation, which is now roughly in balance with load. FERC rejected the proposed process and directed NEPOOL to develop a realistic evaluation process and to develop a system that uses locational prices to manage congestion by March 31, 1999 (FERC 1998b). FERC delayed ruling on the cost-allocation method until it sees the revised proposal.

FERC’s rejection of the NEPOOL proposal suggests there is no way to avoid centralized transmission planning and assessment, based in part on realistic judgments concerning future generation-market conditions. FERC viewed as unrealistic and anticompetitive NEPOOL’s assumptions that any new generation must be able to reach all loads within NEPOOL, that all existing generation remains online (i.e., new generation displaces no existing generation), and that all prior interconnection applications result in construction of new generation.

Texas

The Texas PUC (1998) recognizes the role transmission plays in maintaining reliability and in promoting competitive generation markets. The PUC favors a stronger role for the ERCOT ISO in transmission planning because it can do the work “objectively and impartially from a regional perspective.” Currently, individual utilities prepare transmission plans, which the ISO reviews and evaluates from an ERCOT-wide perspective. The Texas PUC (1999) revised its transmission rule to increase the ISO’s role in transmission planning, stating that the ISO “shall supervise ERCOT transmission system planning and exercise comprehensive authority over the planning of bulk transmission projects . . . .” The PUC is likely to give
considerable deference to ISO recommendations concerning the facilities that need to be built, However, the transmission owners retain the obligation to build new transmission that the ISO determines is needed."

The Texas PUC (1998) stated that “Fundamentally, new transmission investment is a regulated solution.” Congestion pricing is a combination of regulated and market-based solutions that may be used in the future. Texas currently uses a combination of postage-stamp (70%) and distance-based (30%, through a MW-mile charge) pricing.

The ERCOT Technical Advisory Ad Hoc Committee on Transmission Adequacy recognized that the ISO needs guidance from the PUC on policy objectives for transmission performance, including both reliability and market needs (ERCOT ISO 1998). The committee can then establish detailed technical standards for operations and planning. The PUC should have the responsibility to establish appropriate levels of reliability. The committee also recommends use of congestion pricing rather than postage-stamp pricing.

Loads have been growing rapidly in Texas (12% growth in the past two years), and 25,000 MW of new generation (almost half the existing capacity) is planned. In response to this growth, the ISO identified five transmission projects that it believes are needed to alleviate system constraints. It recommended to the Texas PUC that these projects be approved promptly (Jones 1999).

California ISO

California participates in the Western Interconnection transmission-planning process. The ISO supports project ideas from transmission owners (the three California investor-owned utilities), the ISO itself, or any entity that participates in the energy marketplace (California ISO 1998; Miller 1998). The ISO’s objective in having a highly inclusive process is to develop “an ISO Grid that best meets the needs of all its users and maximizes the potential benefits to the State of California. The goal is to meet the reliability needs of the state at the minimum cost to consumers.”

Transmission owners file plans annually for their portion of the grid. These plans are coordinated with neighboring systems and cover a minimum of five years into the future. These plans are required to address the needs identified by other market participants. The ISO reviews all projects to assure that they are properly integrated and meet the ISO’s requirements. The ISO also conducts an economic analysis to determine if the project should be incorporated into the local transmission access fee or split among the participants. The ISO can require the transmission owners to build transmission projects for reliability. The costs of these facilities

---

*The PUC revised its transmission rule for several reasons, including uncertainty among the owners of merchant plants about whether and how much they would have to pay for new transmission facilities, fair treatment for all owners and users of the bulk-power system, and recovery of construction costs incurred by transmission owners.
would then be assigned to a specific transmission owner (i.e., its retail customers) or apportioned among the beneficiaries.

The California ISO (1998), in its report to the state legislature, noted that

… neither existing institutions nor new institutional frameworks being considered provide for the effective integration of reliability, commercial, and planning issues. This can result in new facilities being identified as necessary for reliability, but with no practical means identified of funding the facility addition. … A further issue concerns the need for reliability organizations to coordinate effectively with state and federal regulators who will provide permits for new transmission projects and approve recovery of prudently-incurred costs for reliability compliance in rates. In particular, the cost recovery issue is complicated due to the mix of utilities and non-utilities in the restructured market. Care must be taken to ensure that a level playing field exists so that all market participants have a reasonable opportunity to recover such prudently-incurred costs.

The California ISO is currently considering two competing proposals that would impose different burdens on new generators to relieve transmission congestion (Jubien 1999). The Advance Congestion Cost Mitigation proposal, favored by the ISO and most California market participants, would require a new generator to mitigate any new transmission congestion it would cause as a condition of interconnection to the grid. If connecting a new generator to the grid would cause congestion in that zone, the new generator would be responsible for mitigating (paying for) congestion relief; this approach protects the transmission rights of existing generators. The competing proposal would treat all generators alike and grant no vintage rights to transmission capacity.

Federal Energy Regulatory Commission

FERC (1999) issued a Notice of Proposed Rule Making on RTOs in May 1999 proposing that all its jurisdictional utilities file plans, no later than October 2000, to join an RTO. One of the seven minimum functions of an RTO is to “… be responsible for planning necessary transmission additions and upgrades that will enable it to provide efficient, reliable and non-discriminatory transmission service and coordinate such efforts with the appropriate state authorities.” FERC favors RTOs for transmission planning and expansion because “In a multi-utility region, decisions about where to site new facilities and who should pay for capacity expansions can be … complex unless a regional body provides a forum for discussions and a method for resolving disputes.”

FERC proposes to require the RTO plans to include two key features with respect to transmission planning and expansion. First, the process “must encourage market-driven operating and investment actions for preventing and relieving congestion.” FERC is clear that
it favors market mechanisms to deal with congestion rather than the engineering approach embodied in NERC’s current transmission loading relief procedures. FERC strongly supports market mechanisms to drive transmission investments, encouraging RTOs to propose “… transmission pricing reforms and the explicit use of pricing incentives in encouraging RTOs to make efficient investments in new transmission facilities.”

Second, the process “must accommodate efforts by state regulatory commissions to create multi-state agreements to review and approve new transmission facilities.” FERC believes that the emergence of RTOs might encourage the development of regional regulatory bodies to oversee the certification and siting of new transmission facilities that serve regional, rather than local, needs.

Finally, FERC recognized that a transmission owner that invests in new transmission “may be concerned about recovering its investment.” Therefore, FERC encourages RTOs to propose “… the explicit use of pricing incentives in encouraging RTOs to make efficient investments in new transmission facilities.” Such incentives could include a higher return on equity, accelerated cost recovery, or other mechanisms.

DISCUSSION

Transmission adequacy is a much tougher issue to resolve than is generation adequacy. Much as we might like to rely on competitive markets to make decisions about transmission expansions and their alternatives, the technical characteristics of transmission grids makes this unlikely. These technical limitations include:

- It is rarely possible to directly control the flow of electricity through individual transmission elements.
- The flows across each transmission element affect the flows through almost every other element on the grid.
- Transmission investments exhibit large economies of scale, which makes it much cheaper to add transmission capacity in large amounts rather than incrementally as needed.
- Regulatory approvals for construction of new transmission lines can take several years.
- Transmission investments typically have lifetimes of several decades.
- The high capital costs and low operating costs of transmission ensures that virtually all costs are sunk (and therefore largely irrelevant to markets) once they are incurred.

The physics of transmission requires that many decisions be made by central planners, generally with regulatory approval. But the planners and regulators do not risk their own money; any mistakes they make are paid for by transmission customers.

The transmission-planning processes in place and under development by various ISOs and Transcos allow market participants to offer alternative solutions for the central authority
to choose from. The regional entity could specify the problem in general terms and allow multiple technologies (e.g., generation redispatch, new generation, controllable load, new transmission, or upgrades of existing transmission) to compete to provide the lowest-cost solution. Still, it is the customers and perhaps the transmission owners, not the decision makers, that assume the economic risk that the investment is needed and that its costs will be recovered.

On a positive note, the process can be opened to wide participation. A structure can be developed, as in California, that assures that the planners and decision makers receive input from all parties. By widely disseminating plans for the future of the grid, market participants can better plan their own activities as well as suggest improvements to the transmission plan. Thus, regionalization of the transmission grid, through creation of ISOs and Transcos, helps to resolve transmission-adequacy problems. In addition, congestion pricing provides important economic information to both market participants and central planners concerning the need for and location of transmission enhancements.

The Task Force on Electric System Reliability (1998) noted two problems in deciding whether and where to enhance transmission systems. First, there is little agreement on how to price the use of transmission to create efficient signals concerning new investment in supply or demand. Proposals range from PJM’s nodal pricing to the ISO New England use of postage-stamp pricing. The Task Force encouraged the industry to learn from the different pricing approaches being adopted throughout the country. Second, alternative mechanisms to relieve congestion operate in different market structures, with generation and demand-side options in competitive markets and transmission in regulated-monopoly markets (Exhibit 4).

Joskow (1997) noted several reasons why markets alone may not be able to make appropriate transmission investments: “[they are] lumpy, characterized by economies of scale and can have physical impacts throughout the network … [and] the combination of imperfectly defined property rights, economies of scale, and long-lived sunk costs.” On a more optimistic note, he wrote:

However, there is no reason why the primary initiative for transmission upgrades should not be left to private parties, especially if a reasonably good allocation of capacity rights, whether physical or financial, is created. The network operator could then determine whether proposed upgrades have adverse uncompensated effects on some users of the network, or whether there are inadequate private market incentives for investment because of scale economies or free-riding problems. In those cases the network operator could identify investment projects that the transmission owners would be obligated to build and the associated costs could be recovered from all network users.
Exhibit 4. A Generation and Transmission Parable

Investors can earn or lose money in bulk-power markets for reasons that have nothing to do with the wisdom of their investments. Consider two investors.

Mr G. studies the electricity market, notes trends in load growth and the construction of other generation facilities, and concludes that a new gas-fired combined-cycle unit could lower costs to customers and make money if it was located in Newville, which currently has only high-cost generators and is “downstream” of a congested interface. He invests his money, builds the plant, sells plenty of power, and is doing well. Three years later, the ISO concludes that a transmission line from Newville to Lignite County would eliminate congestion, further reduce power costs in Newville, and generally benefit the transmission grid. Mr. G’s unit goes out of business, he is financially ruined, and he dies a bitter old man in the county home for the poor.

Meanwhile, in another part of the system, Mr. T studies the electricity market and concludes that a new transmission line could lower the cost to customers if it connected Growthtown to Cheapgen. He takes the idea to the ISO, which concurs and obtains regulatory approval. Mr. T builds and operates the line under contract to the ISO. Three years later, Miracle Micro Turbines are installed throughout Growthtown, and the power flows on Mr. T’s transmission line drop to zero. Because the cost of Mr. T’s line is in the transmission rate base, customers continue to pay Mr. T for the line, and he retires on his modest regulated income a happy man.

Investors like Mr. G have to contend with the risk that newer generation or transmission investments will render their investment obsolete. The same is generally not true for investments in transmission, which are protected by regulation. Regulators and regional transmission organizations will have to be careful to ensure that the fear of ending up like Mr. G does not discourage investors from entering the market for new generation.
Generation and transmission adequacy pose troublesome transitional and, perhaps, long-term issues. During the long, awkward, and difficult transition from a highly regulated, retail-monopoly-franchise structure to a competitive structure, adequacy problems arise for both generation and transmission.

Perhaps the key generation-adequacy problem is the absence of a demand-side response to real-time pricing. Economic theory suggests that consumers and suppliers, in response to real-time prices, will take appropriate steps to ensure generation adequacy. But, if most retail consumers continue to face traditional tariff prices that have little or no temporal variation, this approach will be short-circuited. Until real-time pricing is available to at least some retail customers, traditional approaches to maintaining generation adequacy may be needed.

Different problems arise with transmission, centered about the appropriate institutions that will analyze and plan for transmission enhancements, decide on which transmission alternatives should be built, pay for these investments, and recover the costs of these investments from wholesale and retail customers. In the long run, large regional organizations, as recently proposed by FERC, will likely take on most of these responsibilities, but such organizations now exist only in some parts of the country (covering about one-third of U.S. electricity demand), and these ISOs are very new and still evolving. Finally, the importance of congestion pricing and congestion rights, which could expand greatly the role of competition in enlightening transmission-expansion decisions, is still hotly debated and largely untested. Unlike generation, transmission planning and investment will likely continue to be shared between markets and regulators.

Our key findings and conclusions are:

- Generation-capacity margins have been declining for at least a decade and are expected to continue to decline.
- The amount of transmission capacity installed per MW of peak demand has been declining for at least a decade and is expected to continue to decline.
- Transmission-maintenance expenditures have been declining for a decade and may continue to do so.
- Whether these declines in generation and transmission adequacy reflect increased productivity or serious shortfalls in reliability is not clear. It is clear, however, that the transitional state of the U.S. electricity industry (half competitive and half regulated)
leads to tremendous uncertainty, which may limit investments in long-lived assets, such as generating units and transmission facilities.

- Generation adequacy could be maintained in competitive electricity markets in one of two ways: (1) sole reliance on markets, acting through time-varying spot prices or (2) continuation of the historical practice of setting minimum requirements on installed capacity that must be met by all load-serving entities.

- Market-based methods for generation expansion seem, both to us and to most of the people we talked with, the preferred long-term approach.

- Transmission adequacy is a more difficult issue to resolve than generation adequacy because of the difficulty in controlling flows on individual transmission elements, the lumpy nature of transmission investments, and the positive externalities associated with transmission, all of which make sole reliance on markets for transmission investments unlikely.

- Analyses of transmission needs and the construction of new transmission facilities will continue to require central planning and government oversight. These processes will be more complicated than they were in the past because transmission planners will have less timely information on new generation projects and because the planning, approval, and construction process for new transmission lines is a lengthy one. Because RTOs are so new, their “procedures for [transmission] expansion are confusing,” as FERC noted about one ISO but which applies to all the extant proposals. Finally, because new transmission projects will affect the profitability of individual market participants, both generation and load, new transmission projects may be very contentious.

- Agreement is growing that some form of location-based congestion pricing, coupled with physical or financial rights to transmission, is essential in the long term. Such market-based systems will provide signals to investors on where to locate new generation. This price information will also help system planners, regulators, and investors decide when and what transmission investments to undertake.

ACKNOWLEDGMENTS

We thank the Edison Electric Institute for its support of this project. We are especially grateful to our EEI project managers, James Evans and Kenneth Linder, for their ongoing encouragement and guidance. We thank Lawrence Anderson, Marc Butts, James Evans, William Hogan, John Hughes, Sam Jones, Richard Kafka, Pierre Landrieu, Robert Martinko, Pat McAuliffe, Steven Naumann, Richard O’Neill, Armando Perez, Dick Simons, Susan Tierney, Andrew Williams, and Mark Ziering for their very helpful comments on the draft of this report. We thank Fred O’Hara for editing this final report.
REFERENCES


S. Jones 1999, personal communication, ERCOT ISO, Taylor, TX, April 16.


Western Interconnection Coordination Forum 1998, Western Interconnection Biennial Transmission Plan, Salt Lake City, UT, May.
