

ENERGY DIVISION

UNBUNDLING GENERATION AND TRANSMISSION SERVICES
FOR COMPETITIVE ELECTRICITY MARKETS

ERIC HIRST and BRENDAN KIRBY

January 1998

RECEIVED
JUN 17 1998
OSTI

Sponsored by
The National Regulatory Research Institute
Columbus, Ohio

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORPORATION
for the
U.S. DEPARTMENT OF ENERGY
under contract No. DE-AC05-96OR22464

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

CONTENTS

	Page
SUMMARY	v
LIST OF ACRONYMS	vii
1. INTRODUCTION	1
2. DEFINITIONS OF ANCILLARY SERVICES	3
3. CHARACTERISTICS OF ANCILLARY SERVICES	9
CAN THE SERVICE BE UNBUNDLED?	9
SYSTEM-OPERATOR ROLE	10
LOCATION	12
NEEDED FOR RELIABILITY	13
COMPETITIVE PROVISION	14
4. COSTS AND PRICING PRACTICES	17
CHALLENGES	17
INDIVIDUAL SERVICES	19
SUMMARY	26
5. OPERATING-RESERVE BASICS	29
APPLICATION OF OPERATING RESERVES	29
NERC AND REGIONAL DEFINITIONS AND REQUIREMENTS	30
DATA ON GENERATOR OUTAGES	33
6. EMERGING ISSUES ON OPERATING RESERVES	37
TECHNICAL BASIS FOR RESERVE REQUIREMENTS	37
PAYING FOR RESERVES	38
FUTURE DATA NEEDS	38
MIXING FUNCTIONS WITHIN OPERATING RESERVES	39
7. VOLTAGE-CONTROL BASICS	41
TRANSMISSION-SYSTEM VOLTAGE CONTROL	43
VOLTAGE-CONTROL EQUIPMENT	44

8. EMERGING ISSUES ON VOLTAGE CONTROL	51
TECHNICAL GUIDANCE AND COMPENSATION RULES	51
REQUIREMENTS IN A RESTRUCTURED INDUSTRY	53
9. CONCLUSIONS	55
ACKNOWLEDGMENTS	58
REFERENCES	59

SUMMARY

Ancillary services are those functions performed by the equipment and people that generate, control, and transmit electricity in support of the basic services of generating capacity, energy supply, and power delivery. The Federal Energy Regulatory Commission (FERC) defined such services as those "necessary to support the transmission of electric power from seller to purchaser given the obligations of control areas and transmitting utilities within those control areas to maintain reliable operations of the interconnected transmission system."

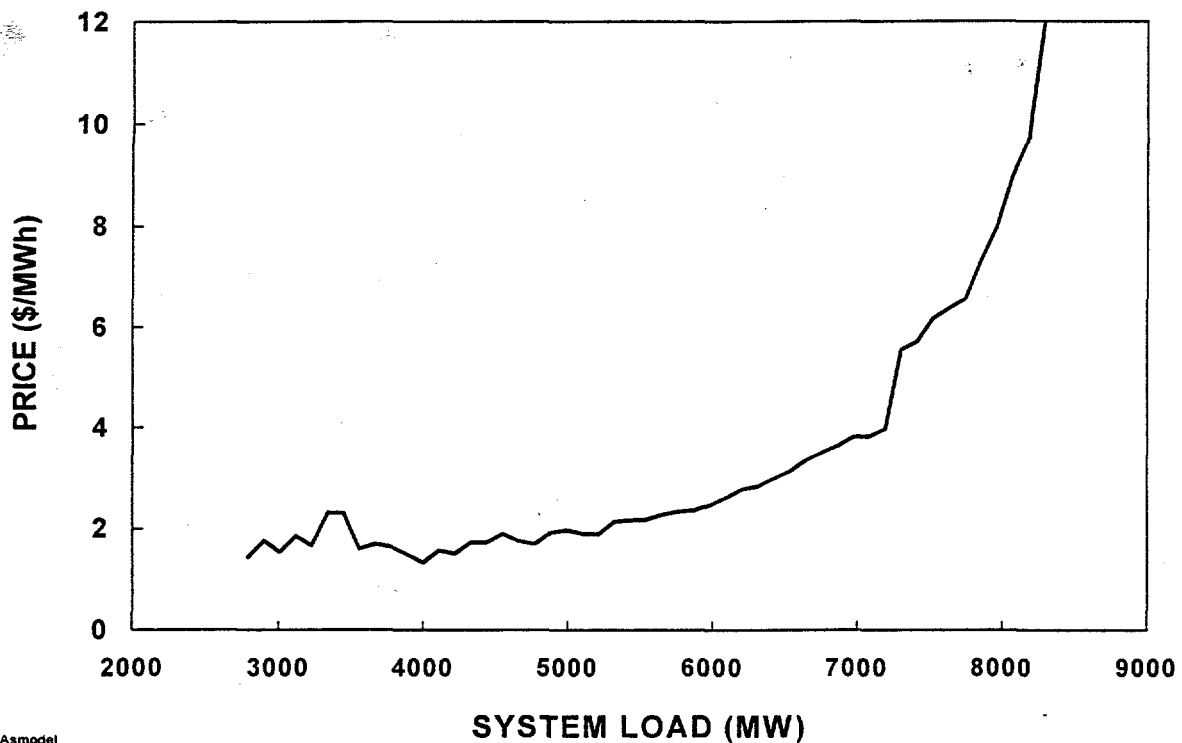
The nationwide cost of ancillary services is about \$12 billion a year, roughly 10% of the cost of the energy commodity. More important than the cost, however, is the necessity of these services for bulk-power reliability and for the support of commercial transactions.

FERC's landmark Order 888 included a *pro forma* tariff with provisions for six key ancillary services. The Interconnected Operations Services Working Group identified another six services that it felt were essential to the operation of bulk-power systems. Several groups throughout the United States have created or are forming independent system operators, which will be responsible for ensuring that appropriate amounts of these ancillary services are available for reliability and commerce.

To date, the electricity industry (including traditional vertically integrated utilities, distribution utilities, power marketers and brokers, customers, and state and federal regulators) has paid insufficient attention to these services. Although the industry has made substantial progress in identifying and defining the key services, much remains to be done to specify methods to measure the production, delivery, and consumption of these services; to identify the costs and cost-allocation factors for these services; and to develop market and operating rules for their provision and pricing.

Developing metrics, determining costs, and setting pricing rules are important because most of these ancillary services are produced by the same pieces of equipment that produce the basic electricity commodity. Thus, the production of energy and ancillary services is highly interactive, sometimes complementary and sometimes competing. In contrast to today's typical time-invariant, embedded-cost prices, competitive prices for ancillary services would vary with system loads and spot prices for energy (Fig. S-1).

The individual ancillary services differ substantially in their features, competitiveness, provision, and pricing. Operating reserves, for example, can likely be provided by competitive markets. The primary supplier cost for this service is the opportunity cost associated with



Asmodel

Fig. S-1. The cost to consumers for eight generation-provided ancillary services as a function of system load for a hypothetical utility system.

foregone energy sales; significant fuel costs are incurred only when these reserves are called upon to respond to the loss of a major generation or transmission outage.

Injection and absorption of reactive power, on the other hand, must be provided close to the location where the voltage control is needed. Therefore, it may not be feasible to create competitive markets for this service. Rather, the pricing and provision of voltage control may continue to be regulated. Capital costs are the dominant costs for this service for both generators and transmission equipment. Opportunity costs arise only when generators are operating at or near full real-power output and are called upon to increase reactive-power output beyond the level associated with the unit's rated power factor. Such operation would require a reduction in the output of real power.

Although FERC has regulatory responsibility for bulk-power markets, state regulators may have a keen interest in these ancillary-service issues. Public utility commissions may choose to get involved in the formation of independent system operators to ensure that suitable levels of reliability are maintained within their states and that market structures promote full and open competition. This report, aimed at state regulatory authorities, provides an overview of the 12 ancillary services plus details on two of those services, operating reserves and voltage support.

LIST OF ACRONYMS

ACE	Area-control error
AEP	American Electric Power Company
AGC	Automatic generation control
ECAR	East Central Area Reliability Coordination Agreement
ERCOT	Electric Reliability Council of Texas
FERC	U.S. Federal Energy Regulatory Commission
ISO	Independent system operator
NERC	North American Electric Reliability Council
NYPP	New York Power Pool
O&M	Operations and maintenance
PJM	Pennsylvania-New Jersey-Maryland Interconnection
PUC	Public utility commission
SERC	Southeastern Electric Reliability Council
SPP	Southwest Power Pool
STATCOM	Static synchronous compensator
SVC	Static VAR compensator

INTRODUCTION

In April 1996, the U.S. Federal Energy Regulatory Commission issued its Order 888 (FERC 1996) to functionally unbundle generation from transmission. This order exposes an array of utility functions commonly called ancillary services. Unbundling does not create these services, but it does require that electricity suppliers, consumers, and regulators consider these functions in ways that were not necessary when they were provided by vertically integrated utilities as part of the bundled electricity product.

Historically, no one other than electric-utility system operators and electrical engineers needed to understand the definitions and applications of these ancillary services. The traditional integrated utilities planned for and deployed generation, transmission, and system-control resources to produce those reliability-related services as part of their normal operations. And the commercial aspects of ancillary services were, until the past few years, unimportant because the vast majority of the wholesale transactions involved vertically integrated utilities that provided their own ancillary services.

The Energy Policy Act of 1992, FERC's Orders 888 and 889, and the growing importance of power marketers and brokers are changing the situation rapidly and dramatically. Groups within different regions are developing proposals for independent system operators (ISOs) and, in some cases, power exchanges. Proposals to create ISOs in California, New England, New York, and the mid-Atlantic region have been filed with FERC. ISO proposals are being developed in the Pacific Northwest and Rocky Mountain region (IndeGo), the Southwest (DesertStar), the Midwest, the Mid-Continent Area Power Pool, and the Southwest Power Pool (ICF Resources 1997). Ancillary services are an integral part of all these proposals because of the vital role that these services play in maintaining reliability and effecting commercial transactions.

This report is aimed primarily at state regulators; it presents the basic concepts of ancillary services to demonstrate their importance for the viability of future electric-industry structure and operation.

While FERC is the primary forum for regulatory adjudication of issues related to bulk-power markets, state regulators can also play a role here. Before opening retail markets to competition, state regulators will want to be sure that traditional levels of reliability will be maintained (Kincheloe and Burns 1997). To this end, states can work with their jurisdictional utilities in the formation of ISOs and can intervene before FERC on these bulk-power issues. The California Public Utilities Commission (PUC) adopted both roles in the formation of the California ISO and power exchange. In some regions, such as the Northwest, state regulators

act as nonvoting members of the groups that are forming ISOs. In certain states, such as New York and California, the electricity markets are sufficiently large that the PUC can exert substantial influence on the form and content of that state's ISO. Some states, such as Florida and Texas, are sufficiently isolated from other states and, as a consequence, can substantially affect ISO development. Finally, state regulators may be able to affect the extent to which their retail customers can participate as suppliers to ancillary-service markets.

The report consists of two parts. The first part (Chapters 2 through 4) identifies, defines, and explains 12 key ancillary services. Chapter 2 begins by defining the key ancillary services that are required to maintain bulk-power reliability, to effect commercial transactions, or both. Chapter 3 then discusses several characteristics of these services, and Chapter 4 discusses their costs and alternative pricing schemes.

The second part of this report, which is more technical, focuses on two disparate services: operating reserves (Chapters 5 and 6) and voltage control (Chapters 7 and 8). Chapter 5 shows how operating reserves are used to recover from a major outage, explains the national and regional requirements for these reserves, and presents data from three large systems on forced outages and the time to recover from these outages. Chapter 6 discusses several key technical and market issues that will need to be resolved during the next few years. Chapter 7 explains the basics of voltage control, shows how and why voltages are maintained within narrow limits throughout transmission grids, and compares the operating and cost characteristics of different types of voltage-control equipment. Chapter 8 discusses several emerging issues that, like the ones discussed in Chapter 6 for operating reserves, must be resolved as the U.S. electricity industry evolves. The final chapter presents key findings and recommendations to state regulators.

DEFINITIONS OF ANCILLARY SERVICES

During the past two years, at least three organizations have developed comprehensive lists and definitions of ancillary services, FERC (1996 and 1997a), the Interconnected Operations Services Working Group (1997), and Oak Ridge National Laboratory (Hirst and Kirby 1996a).^{*} Although the three sets have many similarities in nomenclature, coverage, and definition, they also have differences. Table 1 shows the names and definitions for 12 ancillary services. The seven services that Hirst and Kirby defined cover the six that FERC required transmission providers to offer plus losses. None of these studies detailed the metrics, costs, and appropriate pricing methods for these services; Chapters 3 and 4 deal further with these issues.

FERC specified two services that transmission customers are required to obtain from the transmission provider:[#] system control and voltage control. FERC determined that these services can only be provided by the local control area. FERC specified four additional services that the transmission provider must offer to transmission customers that serve load within the transmission provider's control area: regulation, spinning and supplemental operating reserves, and energy imbalance. For these four services, however, FERC allows transmission customers to obtain the service in any one of three ways: from the transmission provider, from another source, or by self-provision.

FERC also discussed other ancillary services, including real-power-loss replacement, dynamic scheduling, backup supply, and black-start capability. FERC determined that these services either could be provided competitively, were very inexpensive, or were highly location specific and therefore did not require transmission providers to unbundle and offer them as separate services.

^{*}More recently, several entities have filed proposals with FERC to create ISOs, all of which contain descriptions of ancillary services. These filings include those from the Pennsylvania-New Jersey-Maryland (PJM) Interconnection (Atlantic City Electric et al. 1997), New England Power Pool (1996), the New York Power Pool (NYPP 1997), and California (Pacific Gas and Electric et al. 1997). In addition, FERC (1997a), in its Order 888-A, modified slightly its definitions and requirements for ancillary services.

[#]Currently, transmission providers usually own and operate generating units, operate a control area, and are, most often, vertically integrated utilities. Technically, the mere ownership of transmission facilities is not enough to provide and manage ancillary services; only control-area operators have sufficient real-time information to perform these functions. In the future, such functions will increasingly be assigned to independent system operators. *Independent* means an entity that has no commercial interests in electricity markets.

Table 1. Key ancillary services and their definitions

Service	Description	Time scale
Services FERC requires transmission providers to offer and customers to take from the transmission provider		
System control	The control-area operator functions that schedule generation and transactions before the fact and that control some generation in real-time to maintain generation/load balance; Interconnected Operations Services Working Group definition more restricted, with a focus on reliability, not commercial, activities, including generation/load balance, transmission security, and emergency preparedness	Seconds to hours
Reactive supply and voltage control from generation	The injection or absorption of reactive power from generators to maintain transmission-system voltages within required ranges	Seconds
Services FERC requires transmission providers to offer but which customers can take from the transmission provider, buy from third parties, or self-provide		
Regulation	The use of generation equipped with governors and automatic-generation control (AGC) to maintain minute-to-minute generation/load balance within the control area to meet NERC control-performance standards	~1 minute
Operating reserve - spinning	The provision of generating capacity (usually with governors and AGC) that is synchronized to the grid and is unloaded that can respond immediately to correct for generation/load imbalances caused by generation and transmission outages and that is fully available within 10 minutes	Seconds to <10 minutes
Operating reserve - supplemental	The provision of generating capacity and curtailable load used to correct for generation/load imbalances caused by generation and transmission outages and that is fully available within 10 minutes ^a	<10 minutes
Energy imbalance	The use of generation to correct for hourly mismatches between actual and scheduled transactions between suppliers and their customers	Hourly
Services FERC does not require transmission providers to offer		
Load following	The use of generation to meet the hour-to-hour and daily variations in system load	Hours
Backup supply	Generating capacity that can be made fully available within one hour; used to back up operating reserves and for commercial purposes	30 to 60 minutes
Real-power-loss replacement	The use of generation to compensate for the transmission-system losses from generators to loads	Hourly
Dynamic scheduling	Real-time metering, telemetering, and computer software and hardware to electronically transfer some or all of a generator's output or a customer's load from one control area to another	Seconds
System-black-start capability	The ability of a generating unit to go from a shutdown condition to an operating condition without assistance from the electrical grid and then to energize the grid to help other units start after a blackout occurs	When outages occur
Network-stability services	Maintenance and use of special equipment (e.g., power-system stabilizers and dynamic-braking resistors) to maintain a secure transmission system	Cycles

^aUnlike spinning reserve, supplemental reserve is not required to begin responding immediately.

Some entities, especially the vertically integrated utilities that are the primary transmission providers, believe that FERC's requirement that transmission providers offer services but that transmission customers need not purchase these services from the transmission provider is asymmetric and unfair. They fear that when their embedded-cost tariffs are lower than market prices, customers will buy these four services from them. But, when market prices are lower, customers will purchase the services elsewhere, placing them at a competitive disadvantage. The utilities also consider it unfair that only they must continuously stand ready to provide the service to all customers without compensation unless a customer chooses to purchase the service from them.

Others believe that competitive markets will quickly develop for these four services and that embedded-cost tariffs will then be replaced by market-determined prices. Some believe that even voltage control will be amenable to market provision in many cases. Ultimately, as FERC permits market-based pricing for energy, it will likely do so for many of the generator-provided ancillary services. Chapter 4 discusses these issues further.

The Interconnected Operations Services Working Group identified two services that FERC did not discuss at all: load following [also emphasized by Hirst and Kirby (1996b)] and network-stability service. Other entities identified additional services. To illustrate, the New York Power Pool (1997) includes stormwatch as an ancillary service. This service entails the redispatch of generating units to provide additional transmission reserves in anticipation of potential weather-related contingencies. The Electric Reliability Council of Texas (ERCOT) includes responsive reserve, the portion of spinning reserve that responds to outages "within the first few minutes." It also divides energy imbalance into two services, generation-schedule and load-schedule imbalances.

Each service is briefly defined below. Readers interested in additional detail are urged to read the documentation provided by FERC, the Interconnected Operations Services Working Group, and Hirst and Kirby.*

- System control consists of the control-area operator functions that (1) schedule generating units, transmission resources, and transactions before the fact and (2) monitor and control transmission resources and some generating units in real time to maintain reliability. While individual market participants can handle many of the scheduling and dispatch functions, including most of the economic risk, the local control-area operator still has a major role for reliability reasons. This service could also include after-the-fact accounting and billing.
- Voltage control is the use of generating and transmission-system equipment to inject or absorb reactive power to maintain voltages on the transmission system within required ranges (Kirby and Hirst 1997). FERC decided that the costs of voltage control provided

*In principle, loads can provide many of these services, although today loads are permitted to supply only supplemental reserve.

by transmission equipment [e.g., through capacitors, tap-changing transformers, condensers, reactors, and static VAR compensators (SVCs)] should be incorporated into the basic transmission tariffs, and not charged for separately. FERC decided that voltage control provided by generators should be a separate service. (In general, generators can change their production and absorption of reactive power much more rapidly than can transmission-related voltage-control equipment.) Because reactive-power losses are much greater than real-power losses, voltage-control equipment must generally be dispersed throughout the system and located close to where the voltage support is needed.

- Regulation is the use of online generating units that are equipped with governors and AGC and that can change output quickly (MW/minute) to track the moment-to-moment fluctuations in customer loads and unintended fluctuations in generation (Hirst and Kirby 1996b). In so doing, regulation (along with spinning reserve) helps to maintain interconnection frequency, minimize differences between actual and scheduled power flows between control areas, and match generation to load within the control area. This service can be provided by any appropriately equipped generator that is connected to the grid and electrically close enough to the local control area that physical and economic transmission limitations do not prevent the importation of this power.*
- Load following is the use of online generation equipment to track the interhour changes in customer loads (Hirst and Kirby 1996b). Unlike the minute-to-minute fluctuations, which are generally uncorrelated among customers, the hourly and diurnal changes in customer loads are generally correlated with each other.
- Spinning reserve is the use of generating equipment that is online and synchronized to the grid that can begin to increase output immediately in response to changes in interconnection frequency and that can be fully available within 10 minutes to correct for generation/load imbalances caused by generation or transmission outages (Hirst and Kirby 1997c).[#] In principle, loads under the control of the system operator could help provide this service.
- Supplemental reserve is the use of generating equipment and interruptible load that can be fully available within 10 minutes to correct for generation/load imbalances caused

*For all services, electrical proximity to the local control area implies that the cost savings associated with use of a remote generator are not offset by the costs of transmission and transmission-system losses. This concept of electrical proximity applies to regulation, load following, spinning reserve, supplemental reserve, backup supply, loss replacement, and energy imbalance.

[#]The generating units that provide spinning reserve are often the same units that provide regulation. Almost all generators that supply spinning reserve are equipped with governors and AGC.

by generation or transmission outages (Hirst and Kirby 1997c).^{*} Supplemental reserve differs from spinning reserve only in that supplemental reserve need not begin responding to an outage *immediately*. This service may also include the provision of additional generating capacity that must be fully available within 30 or 60 minutes (the exact time depends on the rules of the regional reliability council) and can then be maintained until commercial arrangements can be made (e.g., for two hours) to "back up" the normal supply for the load.

- Backup supply is a service that customers would purchase to protect against forced outages at the generating units that provide their energy or against loss of transmission between their normal supply and their load. This service is used to restore operating reserves to normal levels and therefore must be available within 30 to 120 minutes. Unlike spinning and supplemental operating reserves, which are system services required for reliability, backup supply is a commercial service that supports individual transactions. A customer that did not purchase backup supply would be required to disconnect its load from the grid when its power supply failed or to buy power on the spot market.
- Energy imbalance is the use of generating equipment and fuel to match any differences between actual and scheduled transactions between suppliers and their customers. Energy imbalance is usually measured at hourly increments.[#] FERC's definition of this service includes only load imbalances because it believes that generators can and will meet their schedules accurately. The Interconnected Operations Services Working Group, however, includes both load and generator imbalances in its definition.
- Real-power-loss replacement is the use of generating equipment and fuel to compensate for the transmission-system losses associated with power flows from generators to

^{*}Different reliability regions define different time periods for operating reserves and backup supply. For example, the California ISO requires that operating reserves must be capable of being maintained for two hours, and that what they call replacement reserve must be fully available within 60 minutes and can then be maintained for two hours. The New England Power Pool (1996) proposal calls for a 30-minute operating reserve rather than California's 60-minute reserve. The FERC definition of supplemental reserve is ambiguous. It could mean (1) fast reserves required to respond to a contingency or (2) slower reserves that are used to back up operating reserves and protect against a second contingency.

[#]Three elements are implicit in FERC's definition of energy imbalance. The first includes only the discrepancies within a defined deadband measured over a defined time interval (e.g., $\pm 1.5\%$ for 60 minutes). Backup supply, the second element, would be contractually arranged beforehand between the customer and a supplier (not necessarily the local control area). Unauthorized use, the third element, would result in penalty charges imposed by the local control area if (1) the customer's load fell outside the deadband and (2) the customer had not arranged for backup service or the backup service failed to perform. Because unauthorized use is not a service, its charge need not be based on costs. Rather, its price would be designed to encourage customers to obtain backup service and to discourage them from leaning on the local control area. Alternatively, energy imbalance can be settled at the hourly spot price, eliminating the need for deadbands, penalties, and intricate accounting. The California ISO adopted this market approach.

customers. This service clearly can be provided by a third party because it is so closely related to the basic energy and capacity services provided by generators. (Some people consider losses a part of the basic energy service, not a separate function.)

- Dynamic scheduling is the electronic transfer from one control area to another of the time-varying electricity consumption of a load or the time-varying electricity production from a generator (Hirst and Kirby 1997a). Dynamic scheduling involves metering the power consumption of the load or production of the generator at the same frequency as the other parameters critical to control of the power system, usually every 2 to 8 seconds. The metered signal is sent to the control centers of both the physical-host control area and the electronic-host control area (the control area assuming responsibility for the load or generation). The physical host uses the signal to remove the load or generation from its responsibility, while the electronic host uses the signal to assume control responsibility. This service will help enable the creation of competitive markets for ancillary services by increasing choices for suppliers and customers. This service requires the full involvement of, and coordination between, the physical-host and electronic-host control areas.
- System-black-start capability is the ability of an electrical system to go from an outage or islanding situation to one that synchronizes and reconnects the local system to the interconnection and then restores service to customers. This capability requires backup communication systems between the control center and generating units as well as key transmission equipment; trained transmission personnel who can check the status of substations, switches, breakers, and other transmission equipment; and generating units that can go from a shutdown condition to an operating condition without support from the electrical system. All these capabilities, coordinated by the system operator, are essential during large-scale blackouts and islanding because only if the efforts are coordinated can these black-start generating units start themselves and then produce power that can be used to energize the grid and provide power to start other generating units. This service is, like voltage control, somewhat location dependent, which may limit the ability of third parties to provide the service.
- Network stability is the use of special equipment at a power plant (such as power-system stabilizers or dynamic-braking resistors) or on the transmission system (such as DC lines, FACTS devices,* or energy-storage devices) to help maintain transmission-system stability and reliability. This service is, like voltage control and system-black-start capability, somewhat location dependent.

*FACTS refers to flexible AC transmission systems, the use of high-speed solid-state technologies to control transmission equipment, thereby improving reliability and increasing capacity.

CHARACTERISTICS OF ANCILLARY SERVICES

Building on the definitions of the ancillary services presented in the preceding chapter, this chapter discusses certain characteristics of these services. In particular, it examines (1) whether the service can be unbundled to suppliers and customers, (2) whether the system operator should provide or control delivery of the service, (3) whether the service must be provided from inside the local control area, (4) whether the service is needed to maintain or restore bulk-power reliability (vs whether the services is needed primarily for commercial, transactional purposes), and (5) whether the service can be provided competitively.

CAN THE SERVICE BE UNBUNDLED?

As the column of "Yes" responses in Table 2 suggests, all the services can be unbundled to suppliers. If a service was not unbundled to suppliers, there would be no way to pay suppliers for their provision of the service. In such a case, suppliers would have no incentive to provide the service. Thus, each of the 12 services must be unbundled to suppliers.

The situation with respect to customers is different, however. While many of the services can, and should, be unbundled to customers, not all fit into this category. In principle, the system operator could calculate the system voltage-control requirements associated with each transaction, in much the same way that it could calculate the real-power-loss requirements for each transaction. However, the need to have reactive-power reserves to protect against contingencies makes it difficult to identify customer-specific use of this service. Hourly costs could instead be spread among the system users for each hour. Customer differentiation would then be limited to time of use.

System-black-start capability and network-stability services are, with respect to customer unbundling, like voltage control. Because it is so difficult to identify any transaction-specific component for what are essentially system-reliability services, the costs of these two services should probably also be part of the basic transmission-access charge. Here, however, there is no time differentiation in the costs of the services, so there is no need to have time-varying charges for the service.

The two operating-reserve services, spinning and supplemental, are intended to protect the bulk-power system, not individual transactions, against problems associated with unanticipated generation and transmission outages. As discussed later, the cost of these services varies from hour to hour, so the customer charges should vary from hour to hour. Customers

that require less reliability are free to sell resources to the system in the form of interruptible load (i.e., as supplemental reserve).

System control presents an interesting example with respect to customer unbundling. The system-control portion of this service cannot be readily assigned to customers. However, the scheduling part can be explicitly assigned to individual transactions.

The remaining six services—regulation, load following, backup supply, energy imbalance, loss replacement, and dynamic scheduling—can be unbundled to customers.

SYSTEM-OPERATOR ROLE

By definition, the system operator is the only entity that can provide the system control and scheduling service. However, almost by definition, the system operator cannot provide any of the other services. To the extent that the system operator is *independent* of the owners of generation and transmission, it will own no generating and perhaps no transmission assets. Thus, it will not be able to physically provide these ancillary services.

However, the system operator may need to control the provision of many of these services. The role of the system operator is crucial because it is much more cost effective to provide many of the services (e.g., regulation, voltage control, and operating reserves) for the aggregate load than for each load separately. Indeed, it is desirable to provide the services to individual customers for only a few services, including backup supply, energy imbalance, dynamic scheduling, and perhaps load following and losses.

With respect to system services, the system operator—and only the system operator—knows what the regulation requirements are for the control area from second to second. The operator's knowledge of area-control error (ACE), calculated every two to four seconds, is the basis for its decisions on whether and how to use the regulating margin at its disposal. Thus, although the generators that provide this service are neither owned nor operated by the system operator, their provision of the regulation service is controlled by the AGC signals that the system operator sends to each generating unit that is providing the service.

In a similar fashion, only the system operator, based on its knowledge of power flows and possible contingencies, can set the voltage schedules and reactive-power reserves throughout the transmission grid. During emergency conditions, only the system operator can determine the appropriate voltages throughout the grid and therefore is the only entity that can direct generator reactive-power operations. Therefore, voltage schedules and the resulting reactive-power injection and absorption must be under the control of the system operator.

Table 2. Characteristics of key ancillary services

	Unbundled to		System operator		Inside control area	Needed for reliability	Provided competitively	Can competition determine	
	Suppliers	Customers	Provision	Control ^a				Amount and features	Prices
System control	Yes	Yes	Yes	Yes	No	Yes	No	No ^b	No
Voltage control	Yes	Probably not	No	Yes	Yes	Yes	Perhaps	No ^b	Perhaps
Regulation	Yes	Yes	No	Yes	Generator	Yes	Yes	No ^b	Yes
Load following	Yes	Yes	No	Needs data	must be close	No	Yes	Yes	Yes
Spinning reserve	Yes	Probably	No	Yes	enough to control	Yes	Yes	No ^b	Yes
Supplemental reserve	Yes	Probably	No	Yes	area that	Yes	Yes	No ^b	Yes
Backup supply	Yes	Yes	No	Needs data	transmission constraints	No	Yes	Yes	Yes
Energy imbalance	Yes	Yes	No	Yes	do not apply	Yes	Yes	No ^b	Yes
Real-power-loss replacement	Yes	Yes	No	Yes		Yes	Yes	No ^b	Yes
Dynamic scheduling	Yes	Yes	In part	Needs data	Not applicable	No	No	Yes	In part
System-black-start capability	Yes	No	In part	Yes	Yes	Yes	Perhaps	No	Perhaps
Network stability	Yes	No	No	Yes	Yes	Yes	Perhaps	No	Perhaps

^aSome control is automatic and occurs without any explicit operator intervention (e.g., the use of network-stability equipment or generator governors). Some control is automatic and involves computer-generated signals that are sent from the control center to individual generators (e.g., regulation). And some control is manual and involves human communication (e.g., a telephone call from the system operator to the plant operator to call for supplemental operating reserve).

^bWhile only the system operator can determine how much of a reliability service the system requires at any given time, customers can affect that amount, perhaps in response to price signals, by changing the time of consumption, the source of supply, or the characteristics of their load. In that sense, the service quantity can be at least partially determined by the market.

Analogous situations apply to the operating-reserve services, energy imbalance, black-start capability, and network-stability services. For all of these services, the system operator is the only entity with sufficient and timely information to decide how much of each service is required. In addition, system provision of the service, rather than customer provision, provides economies of scale. That is, fewer resources are required to provide a given level of service to an aggregation of loads than to the sum of the individual loads.

The operator does not need to control the provision of backup supply or dynamic scheduling. However, it needs up-to-date (i.e., once every several seconds for dynamic scheduling) information on the status of these services and their provision. This information requirement is a consequence of the system operator's responsibility to maintain generation/load balance within the control area.

The operator does need to control the real-time provision of losses because only it knows what the losses are in real time. But the operator does not need to choose the generators that provide for loss compensation.

LOCATION

Three of the services—voltage control, system black start, and network stability—generally must be provided from within the control area. These services must be delivered to the transmission grid at locations that only the system operator can determine. The electrical "location" and not the location relative to the control-area boundaries determines where these services are required. In Order 888-A, FERC (1997a) modified its earlier requirement that generators supplying voltage support be located inside the local control area. FERC now requires only that the generation be "under the control of the control area operator" FERC, however, will leave it to the control-area operator to decide whether the generator is suitably located to provide this service.

System control can be provided from any location that can maintain secure telecommunications contact with the equipment that monitors and controls loads, generation, and tieline flows for the control area. This service is provided from the system operator's control center, which need not be located within the control area. For example, both the Western Area Power Administration and the Salt River Project have their control centers in Phoenix, which is inside the Arizona Public Service Company control area.

Seven services—regulation, load following, spinning reserve, supplemental reserve, backup supply, energy imbalance, and loss replacement—can, in principle, be provided from any location within the interconnection.* In practice, the generating units that provide these services must be located close enough to the control area so that transmission charges and

*If provision of the service requires AGC, then its delivery from outside the control area requires the use of dynamic scheduling.

losses do not cost too much. In addition, the generators must be located so that they are unlikely to be on the wrong side of a transmission constraint. For example, if a generator that provides spinning reserve is located on the far side of a failed transmission line and if the failure of that line limits the transfer capability of the system, then the output from that generator may not be able to be delivered to the control area to provide the service. The system operator should have the final say on whether a generator is in a location that allows it to deliver *reliably* the particular service. In making such determinations, the system operator will consider contingencies (i.e., the potential failure of various generators and transmission elements) as well as normal operating conditions.

Because dynamic scheduling involves the electronic transfer of generation or load, at least part of the service must reside within the physical-host control area.

NEEDED FOR RELIABILITY

The extent to which services are needed to maintain bulk-power reliability vs to enable commercial transactions is difficult to identify unambiguously. In many cases, the service is needed for both purposes. A reliable electric system can be defined as one that allows for few interruptions of service to customers. Outages can be defined in terms of their number, frequency, duration, and amount of load (or number of customers) affected. Equally important, but much more difficult to quantify, is the value of loss of load.*

Bulk-power systems are fundamentally different from other large infrastructure systems, such as air-traffic-control centers, natural-gas pipelines, and long-distance telephone networks. Electric systems have two unique characteristics:

- They need continuous and near instantaneous balancing of generation and load, consistent with transmission-network constraints; this requires metering, computing, telecommunications, and control equipment to monitor loads, generation, and the transmission system and to adjust generation output to match load.
- The transmission network is primarily passive, with few "control valves" or "booster pumps" (e.g., phase-angle regulators, FACTS devices, and DC lines) to regulate electrical flows on individual lines; control actions are limited primarily to adjusting generation output and to opening and closing switches to reconfigure the network.

These two unique characteristics lead to four reliability consequences with practical implications that dominate power-system design and operations:

*A 10-minute power outage in a residence is an annoyance because someone has to reset the digital clocks but imposes only small economic costs. But a similar outage for a computer-chip manufacturer might entail the loss of millions of dollars of output.

- Every action can affect activities at remote locations on the grid. Therefore, the activities of all participants must be coordinated.
- Cascading problems that increase in severity are a real problem. Failure of a single element can, if not managed properly, cause the subsequent rapid failure of many additional elements, disrupting the entire transmission system.
- The need to be ready for the next contingency, more than current conditions, dominates the design and operation of the bulk-power system. Usually, the present flow through a line or transformer does not limit allowable power transfers, but rather the flow that would occur if another element fails.
- Because electricity flows at near the speed of light, maintaining reliability often requires that actions be taken almost instantaneously (within a fraction of a second), which requires computing, communication, and control actions that are automatic.

The reliability requirement to maintain a near-instantaneous balance between aggregated generation and system load results in classifying regulation, the operating reserves, real-power-loss replacement, and energy imbalance as reliability services. Voltage control and network stability are also reliability services because they are instantaneous in nature, and without them the network cannot function. System black start is a reliability service because without it the network could not recover from a collapse. Similarly, system control is a reliability service because the network cannot function without it.

Load following is not classified as a reliability service for two reasons. First, energy imbalance will take care of any discrepancy between load and generation if load following is not provided. (Alternatively, load following could be classified as a reliability service, and energy imbalance could be optional.) Second, load following bridges the short times where markets are not effective and the longer times when markets work well. Similarly, backup supply is not required for reliability because there is ample time for a customer to arrange for this service. Failure to arrange for the service would result in the customer going to the spot market for the service or disconnecting its load from the grid. Dynamic scheduling is an optional service that helps expand commercial opportunities.

COMPETITIVE PROVISION

If a service can be provided competitively, then FERC may be able to drop its requirement that transmission providers offer the service under FERC-approved rates. Over time, competitive markets will likely develop for certain services. If a service can be provided competitively, one can then ask whether markets can determine (1) how much and the features of the service to provide and (2) the prices charged for the service.

Competitive markets can almost surely be created for seven of the generator-related services, including regulation, load following, both operating reserves, backup supply, energy imbalance, and real-power-loss replacement. That is, markets can determine the suppliers and prices for these services. Any generator connected to the grid and electrically close to the local control area can provide these seven services.

Competitive markets could develop for three other services: voltage control, black-start capability, and network stability. To maintain reliability, these services must be delivered at specific locations within the control area. Only the system operator can determine where these services should be located. For example, it would make no sense to have all generators with black-start capability located in one corner of the control area. Consider what would happen if that portion of the control area was, because of a major disturbance, in a different island or on the other side of a major transmission failure than the rest of the control area. In that case, those generators would not be able to energize the transmission grid and deliver power to other generators in the rest of the control area.

However, if there are a few generators located so that they can provide the particular service, the competition among them may be enough to allow markets to determine the prices of these services. If the costs of providing the services are low enough (i.e., the barriers to entry are low), the "threat" that other potential suppliers will install the equipment and offer the service may be enough to create a competitive market.

System control and dynamic scheduling are the only two services that unambiguously cannot be provided competitively. These two services, by definition, can be provided only by control centers that manage control-area operations. Although dynamic scheduling must involve the physical and electronic control areas, its use can expand the geographic size of competitive markets for several services. That is, dynamic scheduling makes it possible to acquire some services from generators located outside the local control area. Because only one provider exists for system control and key elements of dynamic scheduling, these services must remain regulated.

Those services needed to maintain bulk-power reliability (discussed above) are required by the system as a whole. Only the system operator has the information required to make decisions concerning the amount of service that the system requires at each time. Still, the aggregate load is composed of individual loads and individual transactions, and individual changes affect the aggregate. Competitive markets could operate if customers are willing and able to change behavior in response to price signals. If the market price for regulation is high during peak hours, for example, a customer could either shift its time of consumption or reduce the fluctuations in its load and thereby affect the system's regulation requirement and the resulting price. Similarly, a customer could affect the system cost for voltage control by either changing its time of consumption or changing its generation source. To effect this change, the price the customer sees for each service must reflect the amount of service each customer takes and the burden the customer places on the system. These relationships will be nonlinear for

most reliability services. Competitive markets can more easily determine the amounts of the services that provide an exclusively commercial function. These services include load following, backup supply, and dynamic scheduling.

On the other hand, competitive markets could play more straightforward roles in determining the suppliers and prices of most of the services. The location-specific requirements for voltage control, black start, and network stability complicate market pricing for these services. It is likely that these services will continue to be priced on an embedded-cost basis, overseen by FERC, for longer than is true for the other services. Ultimately, even these three services may be amenable to provision and pricing by competitive markets.

COSTS AND PRICING PRACTICES

CHALLENGES

Identifying the specific factors that determine the costs for these 12 services is difficult for two reasons:

- Many of the services are provided by equipment that provides other services, and disentangling the contribution of specific components to specific services is difficult. Provision of one service may limit provision of another service. For example, the capacity of a generating unit that is assigned to operating reserves cannot be used to produce energy. Alternatively, provision of a service may require that the unit be operating at a minimum real-power level. For example, a generator must be operating above its minimum output level to provide regulation service (which requires that the generator be able to reduce as well as to increase output to follow moment-to-moment fluctuations in system load). In either case, the primary cost of the ancillary service may be associated with forgoing energy sales or the operating cost associated with minimum energy production.

There are three implications of the joint-production feature of several ancillary services. First, these relationships between provision of ancillary services and the competitive position of the unit in the energy market (i.e., its variable cost relative to that of the marginal unit that hour) affect the hour-by-hour costs of the ancillary service. Second, these relationships imply an opportunity cost to the generator for provision of several services.* That is, if a unit commits to provide a certain service, then it will not be able to provide another service or sell energy into the spot market. Third, these relationships can increase the cost of energy to the system: if a unit commits to provide a certain service, then it will not be able to provide energy into the spot market, and another, more expensive unit will be called upon to make up this deficit, driving up the spot price of energy.

*The opportunity cost represents the highest-valued use of a particular piece of equipment or process. If a generator can make more money selling reserves than it can selling energy, it will sell reserves. Supplier decisions on how to allocate production among services will depend on the costs to produce each service at the particular facility relative to the current price for each service (the difference representing profits).

- These ancillary services have been produced all along by traditional utilities as part of the bundled electricity product they provide to their customers. Because the utilities sold them as part of a bundled product, even they have only limited knowledge of the actual costs to produce each service.

Traditionally, utilities viewed ancillary services as part of overall system requirements. They translated these reliability and other requirements into design and operations specifications for equipment throughout the electrical system. Part of the overall design specification for a 1000-MW power plant, for example, might require that it be capable of maintaining rated real-power output while delivering 480 MVAR for voltage support (equivalent to a 90% power factor). The overall design of the power plant would then be optimized to meet all of the technical specifications (including voltage control) at minimum total cost. It is very difficult, after the fact, to unambiguously determine the costs attributable to voltage control separate from the costs attributable to real-power production. Determining the costs for the other ancillary services presents similar problems.

Further, system operators use a pool of resources to provide these services and seek to minimize the total cost of operating the entire system while remaining within the constraints imposed by reliability requirements (maintaining a specified spinning reserve, for example). Although they may use sophisticated analytical tools to achieve an optimal dispatch, they are not able to determine the individual cost of each service.

Also, utility planning historically focused on overall system costs, with little explicit attention to the costs of *individual* pieces of equipment. For example, the decision to include a particular generating unit in a utility's resource plan would be based on two analyses of the total costs of running the entire system, one with the resource in question and the second without that resource. Use of this aggregate approach means that utilities were not particularly sensitive to tradeoffs between individual resources that might have little impact on overall system costs but might dramatically affect decisions on the individual resources in question.

These factors suggest that the FERC-filed tariffs for ancillary services may bear little relationship to actual costs. One benefit of market provision of the services is that the tradeoffs and the associated risk are shifted from regulated transmission providers to unregulated and competitive generation markets. The flexibility inherent in markets can be used to rapidly adjust prices as experience is gained. Still, the prices charged should approach the underlying costs. If an entity offers prices that are too high, other providers will offer lower prices and gain market share. If an entity offers prices that are too low, it will lose money and go out of business. [Hirst and Kirby (1997b) developed a simple model to simulate the operation of competitive markets for the generator-related ancillary services.]

As discussed above, several of the ancillary services exist primarily to support the entire bulk-power system and cannot be readily assigned to individual transactions or individual customers. These services include system control, black-start capability, and network-stability. Probably the most equitable way to set the customer charges for these services is on the basis of the customer's peak monthly or annual demand (e.g., $\$/MW_{\text{peak}}\text{-month}$). Alternatively, these charges could be based on customer consumption (e.g., $\$/MWh$). These charges could be included in the customer's transmission-system access charge.

INDIVIDUAL SERVICES

Given these considerations, the factors that determine the costs for each service include the following:

System control involves the purchase, maintenance, and use of metering, telemetering, communication, and computing equipment, as well as the people who manage these systems. Compared to the cost of generation, system control is very inexpensive. Splitting the cost of system control into its components is likely to be difficult. These components could include some or all of the activities listed in Table 3. Some of these services, in particular unit commitment, maintenance scheduling, fuel planning, generation planning, and possibly economic dispatch, will likely be performed by other entities operating in competitive markets for electricity.

One area where subdivision of control-center costs may be worthwhile is the distinction between reliability-related functions and support of individual transactions. If such a cost split is feasible, individual suppliers and customers will pay for all the costs associated with their schedules and changes in schedules. Even here, however, it will be difficult to decide whether to assign these charges on the basis of average or marginal costs. The transaction-specific elements should be charged on a per schedule basis (e.g., $\$/\text{schedule}$ and $\$/\text{schedule-change}$).

Voltage control involves the installation and use of equipment at generating units and throughout the transmission grid to maintain voltages within required limits. The costs of such transmission-system equipment (including capacitors, tap-changing transformers, and SVCs) can be readily determined and assigned to this service. FERC has determined that these costs should be included in the basic transmission charge. [See Chapters 7 and 8 as well as Kirby and Hirst (1997) for details on voltage control.]

The generation-related costs of voltage control, which FERC decided should be treated separately, are difficult to isolate. These costs are primarily associated with the capital costs of equipment, especially exciters and automatic voltage regulators. However, these pieces of equipment are required for the production of real power as well as reactive power, which complicates the assignment of costs to each service. In principle, one could determine the cost of a generating unit if it had zero reactive-power capability. The incremental cost associated with the ability to produce and absorb reactive power could then be assigned to voltage control.

Additional capital costs are associated with the increased capacity required for stepup transformers (e.g., for a 1000-MW generating station with a rated power factor of 90%, the transformer rating must be increased to 1100 MVA).

Table 3. Services that might be included in system control

Service	Time scale	Description
Automatic protection	Instantaneous	Minimize damage to equipment and service interruptions caused by faults and equipment failures
Disturbance response	Instantaneous to minutes to hours	Adjust generation and transmission equipment to restore system to scheduled frequency and generation/load balance as quickly and safely as possible
Control of regulation	Seconds to minutes	Adjust generation to match scheduled intertie flows and system load
Control of voltage	Seconds to minutes	Adjust generation and transmission resources to maintain system voltages
Economic dispatch	Minutes to hours	Adjust committed units to maintain frequency and the generation/load area-interchange balance at minimum cost subject to transmission, voltage, and reserve-margin constraints
Unit commitment	Hour ahead to week ahead	Decide when to start up and shut down generating units, respecting unit ramp-up and -down rates and minimum runtimes and loadings
Maintenance scheduling	Up to 2 years	Schedule and coordinate interutility sales and planned generating-unit and transmission-equipment maintenance to maintain reliability and to minimize cost
Transmission planning	1 to 5 years	Design regional and local system additions to maintain reliability and to minimize cost
Fuel planning	1 to 5 years	Develop least-cost fuel supplies, contracts, and delivery schedules
Generation planning	2 to 10 years	Develop a least-cost mix of new generating units, retirements, life extensions, and repowering based on long-term load forecasts

In addition to the capital cost discussed above, provision of reactive support may involve an opportunity cost. This opportunity cost refers to the tradeoff between a generator's ability to produce real vs reactive power. At power factors close to 1, reactive-power production results in almost no loss in real-power-production capability. But as the power factor drops, the real-power penalty increases. If a generator would otherwise be operating at or near its peak

real-power output, a requirement to increase VAR output would necessitate a reduction in real-power output. The loss of revenue from these foregone energy sales is an opportunity cost. Conversely, if a unit is needed to provide reactive support at a time when it would otherwise not be operating at all (because its variable cost is higher than the current spot price), all of the operations, maintenance, and fuel costs, less the price received for the energy produced, are attributable to voltage control. These energy-related costs can make the price of voltage control quite volatile.

Because voltage control serves primarily to support the entire bulk-power system rather than individual transactions, it might be difficult to identify the voltage-control burden created by each transaction.* But the cost to provide voltage control to the system varies from hour to hour, so it is reasonable and beneficial to charge for the service on an hourly basis. Such pricing would provide customers the correct economic signal to help them schedule consumption. The total cost to the system for voltage control for the hour could be allocated among customers based on their demand for the hour (e.g. \$/MW-hour).

Generators that supply reactive power could be paid on the basis of their hourly contracted capability (e.g., in \$/MVAR-hour). If enough appropriately located generators are available to allow a competitive market to operate, the generators' bids would include their opportunity costs. If there are not enough generators and the service must remain regulated, the compensation should include a calculation of the opportunity costs based on the hourly spot-market price and the unit's marginal production cost.

Regulation involves the use of generators and fuel to follow minute-to-minute fluctuations in system load. The frequent changes in output at a generator that provides regulation increases costs in six ways.

1. These output changes raise heat rates, which increases fuel costs.
2. The frequent changes in output and in direction increase operation and maintenance (O&M) costs because of the increased wear and tear on the unit.
3. This increased wear and tear reduces the operating life of the unit.
4. As with voltage control, the need to provide regulating headroom imposes an opportunity cost on that generator because it cannot sell all the output it could produce. This opportunity cost may cause a change in the marginal unit and therefore the market price of energy, possibly a hidden cost of regulation.

*In principle, one could calculate the effect of each transaction on the need to inject or absorb reactive power at various points in the grid. Given the highly interactive nature of reactive requirements with all the power flows on the grid and the need to maintain adequate reactive reserves, assigning customer-specific charges for voltage control is probably not worth the effort.

5. Similarly, the need to provide "footroom" may require a unit to produce more energy than it would otherwise, incurring an uneconomical fuel cost and imposing an opportunity cost on other lower-cost units.
6. Automatic control of the unit output from a remote location requires communication and control equipment, imposing additional capital cost.

Although utilities qualitatively understand these factors, they have little data on the actual costs associated with these factors. Therefore, they know neither the overall cost to provide regulation nor how that cost varies with the amount of regulation provided (MW), the ramp rate (MW/minute) of the unit, or the number of sign changes (between increases and then decreases in unit output).

Although regulation is provided to the system in aggregate, that aggregation is composed of individual loads. Because the fluctuations that are important to regulation are generally uncorrelated among individual loads,* the relationship between individual loads and the total is nonlinear although well defined. Still, reductions in individual load fluctuations reduce the total regulating requirement (Hirst and Kirby 1996b).

Pricing for regulation should more accurately reflect costs than do today's methods, which almost always base charges on average load. The three components of regulation are the standard deviation (or some other measure of volatility), the ramp rate (speed), and the number of sign changes associated with the load; metering limitations at most customer sites will likely prevent adoption of such sophisticated pricing. These three statistics could be calculated for each hour based on the 60 1-minute averages. (Generators do not follow the faster fluctuations, which supports the 1-minute average.) Given the lack of data on these components, regulation could, initially at least, be charged to customers on the basis of customer load volatility as measured by the standard deviation. Because the per-unit cost to follow these largely random and uncorrelated fluctuations in load decreases with increases in the number of customers, the actual charges will need to be scaled down to protect against overcollection. The system operator can probably rely on competitive markets to procure regulation. Generator bids would reflect both the physical costs and the opportunity costs for providing the service. The effects of redispatch to account for generation capacity providing regulation should be charged to the regulation service and not reflected in the basic energy price.

Load following is the interhour equivalent of (intrahour) regulation. The amount of capacity needed for load following is up to 10 times as much as needed for regulation but the required ramp rate is much slower. The primary cost of this service is the opportunity cost associated with the need to commit some of the unit's capacity to follow a particular load shape. This commitment reduces the opportunity to sell energy on the spot market. In addition, load following may involve delivery of energy as well as capacity. This energy cost will be

*Generators also contribute to the minute-to-minute system fluctuations.

associated with fuel and variable O&M. As with regulation, the need to withhold capacity from the energy market to provide load following can impact the price of energy as well. The higher correlation among individual customer interhour fluctuations results in a more linear relationship between individual interhour changes and the need to provide load following.

The charge for load following, which will likely be determined by markets rather than a system operator, will probably be based on the amount of capacity assigned to this service each hour (i.e., \$/MW-hour) plus a charge for the energy provided during the hour (\$/MWh). The energy charge is likely to be based on the current spot-market price for energy. Because the capacity charge will be dominated by the opportunity costs, it will also fluctuate with the spot price of energy. In addition, if a load is moving counter to the system load (e.g., increasing from one hour to the next during the afternoon dropoff), it might receive a *credit* for the capacity portion of the load-following cost.

Spinning reserve refers to generation online, synchronized to the grid, not at full output, able to begin responding to an outage immediately, and capable of providing its full response within 10 minutes. Spinning reserve is usually spread among several units, both for reliability reasons and to provide enough speed (in MW/minute). Two factors affect the overall cost of energy. First, some of the units used to provide energy under least-cost dispatch are backed off to provide spinning reserve, requiring the use of more expensive generating units to provide energy. Second, the units that provide spinning reserve are operated with their valves partially closed to permit the valves to open fully when frequency drops and the governor calls for more turbine-generator output; operating with valves partially, rather than fully, open degrades the unit's heat rate and increases its operating costs (Curtice 1997). The cost of the unit's governor and control system, which responds automatically to frequency deviations, can be attributed in part to spinning reserve (with the remainder assigned to regulation). Similarly, part of the cost of AGC equipment can be associated with spinning reserve. As with regulation, load following, and voltage control, an opportunity cost is often associated with keeping capacity available on a unit to respond as spinning reserve. And operating a unit solely to provide spinning reserve incurs additional costs. [See Chapters 5 and 6 as well as Hirst and Kirby (1997c) for additional detail on spinning and supplemental reserves.]

Customers may be allowed to provide this service through automatic load shedding of part or all of their loads. Costs then would be associated with the automatic load-shedding equipment, including communication equipment. The largest cost might be associated with lost or damaged production when the service is required. Customers are most likely to participate in operating-reserve markets if the prices paid for these services reflect the underlying hourly variation in costs (primarily because of temporal changes in opportunity costs) and if they can move into and out of these markets as their costs and benefits change.

Supplemental reserve refers to generation that can respond fully to an outage within 10 minutes. *Replacement reserve* refers to generation that can be made available within 30 or 60 minutes and is then used to back up operating reserves. The primary cost for these reserves is

the opportunity cost of holding some generating capacity off the market and available for emergency use. Because the units providing these reserves need not be operating (as with spinning reserve), their costs would normally be less than that for spinning reserve. If the spinning reserve market clears first, however, it may use the cheapest resources. Supplemental and replacement reserves might then be more expensive, in a competitive market, than spinning reserves.

Loads can provide supplemental and replacement reserves more easily than they can provide spinning reserves through automatic load shedding. Load-shedding costs will likely be lower for supplemental reserves than for spinning reserves because more time is available (up to 10 minutes for supplemental and up to 60 minutes for replacement reserves) for the customer to take protective action before disrupting its process.

The charges for operating reserves, both spinning and supplemental, will reflect primarily the capacity assigned to these services each hour, captured in a \$/MW-hour charge. Because these services are amenable to market provision, the capacity cost will primarily reflect any lost opportunity related to the spot market. When these reserves are used to replace generation that experiences a forced outage, additional fuel plus variable O&M costs will occur. The cost of operating reserves could be assigned to the generators, with the fixed costs allocated to generators on the basis of their demonstrated reliability (e.g., forced-outage frequencies) and the variable costs assigned to the generator that caused the problem (i.e., that experienced the forced outage that required the use of operating reserves). The variable cost would likely be based on the energy bids from the generators that the system operator calls on to provide reserves. The costs of operating reserves could be collected from generators rather than directly from customers to provide an incentive for generators to maintain an appropriate level of reliability. (Of course, ultimately customers pay for the costs of operating reserves.)

Backup supply refers to generation that is available under contract to provide energy when the generator associated with a particular transaction fails. In addition to incurring the same costs as the primary supplier when the service is actually delivered, the supplier incurs opportunity costs associated with being ready to supply this service. These costs are not as severe as those incurred by generators supplying operating reserves because the advance notice for provision of backup supply is typically 30 to 60 minutes. Therefore, the generator supplying this service is free to engage in short-term transactions (e.g., sell nonfirm energy). Costs are further reduced because backup supply can often be simultaneously provided to multiple generators if the probability of simultaneous failures is low.

The charge for backup supply will be determined by markets rather than by a system operator. The charge will likely be split into a reservation charge and an energy charge with the latter assessed only if the backup generation is used. Given its insurance-like qualities, the reservation charge will likely be based on the probability that the backup generation will be called upon and the amount of capacity reserved (not necessarily equal to the customer's

current load), expressed in \$/MW-reserved. The energy charge (\$/MWh) may be either a guaranteed future price or the current spot price.

Energy imbalance is the hourly analog of regulation. Whereas the regulation service provides *capacity* from minute to minute, the energy-imbalance service provides *energy* over the hour. Its costs include both fuel and capital components. The cost factors for energy imbalance are the same as those for the basic energy service. Therefore, the costs and prices for energy imbalance will equal those in the hourly spot market for energy.

Energy imbalance will likely be priced in \$/MWh and charged on the basis of a transaction's *net* hourly imbalance (i.e., the difference between generation and load imbalances) and the current hourly spot-market price. Ultimately, FERC's complicated system of deadbands and repayment in kind will likely disappear and be replaced by market-based systems, as proposed by the California ISO.

Real-power-loss replacement is the real-time provision of energy and capacity to compensate for losses in the transmission system. As with energy imbalance, the cost factors for loss replacement are the same as those for the basic energy service. Because losses vary with total load and prices generally increase with increasing load, the cost of losses increases nonlinearly with system load; marginal losses are roughly double average losses. Compensation for losses, if based on marginal losses, must be adjusted downward as in the California ISO proposal to prevent overcollection of revenues.

Loss replacement will also likely be priced in \$/MWh and charged on the basis of the current hourly spot-market price with the amount of service consumed computed hourly. As with load following, if a transaction is moving power counter to prevailing flows, it may receive a *credit* for losses, because the transaction is reducing overall system losses.

Losses depend on the locations of generators and loads. Some ISO proposals (in particular, California) simplify the calculations and assign losses on the basis of generation location only. Others (especially PJM and New York) assign losses on the basis of locational marginal pricing.

Dynamic scheduling is the electronic transfer of a load or generation from one control area to another. It involves both initial and ongoing costs. Initial costs include those related to the purchase and installation of additional metering, telemetering, communication, and computing equipment. For example, system operators may need to obtain or upgrade communication systems, such as leased telephone lines, microwave systems, or fiber-optic-cable systems. Similarly, energy-management systems in control centers may need to be expanded, involving both hardware and software. In other cases where data are already being collected and telemetered to one control center, initial investments might be quite modest. The primary ongoing costs relate to communication systems (especially for leased telephone lines) and to periodic inspections, maintenance, and repair of field equipment.

The capital and maintenance costs for dynamic scheduling will likely be assessed for each installation. The operating costs could be similar in form to the charges for static scheduling (\$/schedule) but should be lower because less operator interaction is required for each change.

Black-start capability is the ability of an isolated electrical system to start itself without support from the grid. This service, which encompasses generating units, transmission equipment, and the control center, includes both capital and operating costs. The capital costs are for the energy supply that permits black start (e.g., a diesel generator, transformers, and fuel-storage tanks) and for communication and synchronization equipment. The operating costs are of two kinds. First, costs are incurred when the system is tested to be sure that it can provide the service when needed. Such costs could include tests of the local energy source and transmission switching plus operator training. Second, costs are incurred when islanding occurs and the unit has to start itself and energize the grid; these costs include fuel and labor.

Network-stability services refers to the use of equipment located at generating stations to maintain reliability of the grid. The primary costs for these services are capital, testing, and maintenance. Few operating costs are associated with these services.

As noted above, black-start capability and network-stability services support the entire bulk-power system rather than individual transactions. Therefore, their prices could be incorporated within the basic transmission tariff, normalized on the basis of a customer's total electricity consumption or peak demand.

SUMMARY

Table 4 summarizes the preceding discussion concerning the costs associated with these 12 ancillary services. The table also shows the basis on which each service might be priced to consumers. (In most cases, the pricing basis for payments to suppliers would be the same as that used to collect revenues from customers.)

Using data, assumptions, and analyses from 12 utilities throughout the United States, Kirby and Hirst (1996) developed estimates of the costs of ancillary services. In aggregate, these costs ranged from \$1.5 to \$6.8/MWh, with an average of \$4.1/MWh for the 12 utilities in this sample. Losses are the most expensive ancillary service, accounting for 30% of the total (Table 5). Because these estimates are based primarily on embedded costs rather than prices set in competitive markets, they should be viewed primarily as initial estimates of these costs. Curtice (1997) analyzed FERC filings from 20 utilities and found average prices for voltage control of about \$0.39/MWh (close to the Kirby and Hirst estimate) and for spinning reserve of \$4.3/MWh (much higher than the Kirby and Hirst estimate).

Hirst and Kirby (1997b) analyzed competitive markets for ancillary services. For their base-case system, ancillary-service costs amounted to about 10% of the direct energy costs that

customers experience. Losses were the most expensive service, accounting for about a third of the total ancillary-services bill (Fig. 1).

Table 4. Cost-allocation and pricing basis for 12 ancillary services

Service	Cost-allocation factors				Pricing basis
	Capital	Fuel	O&M	Opportunity	
System control	✓		✓		Basic charge ^a + \$/schedule
Voltage control	✓			✓	Hourly basic charge ^{a,b}
Regulation	✓	✓	✓	✓	Standard deviation, ramp rate, and number of sign changes for 1-minute load, costs vary hourly
Load following	✓	✓	✓	✓	\$/MW _{reserved} -hr for capacity + \$/MWh for energy
Spinning reserve	✓	✓	✓	✓	\$/MW _{demand} -hr + \$/MWh for energy
Supplemental reserve	✓	When used		✓	\$/MW _{demand} -hr + \$/MWh for energy
Backup supply	✓	When used		✓	\$/MW _{reserved} -month
Energy imbalance	✓	✓	✓		\$/MWh, costs vary hourly
Real-power-loss replacement	✓	✓	✓		\$/MWh, costs vary hourly
Dynamic scheduling	✓		✓		\$/schedule or \$/MW _{peak} -month
System-black-start capability	✓	When used	✓		Basic charge ^a
Network stability	✓				Basic charge ^a

^aThe basic charge could be normalized to a customer's monthly peak demand or to a customer's demand at the time of system peak, e.g., \$/MW-month.

^bCompensation for voltage-control reservation could be on the basis of \$/MVAR-hr.

Table 5. Average ancillary-service costs (\$/MWh) for 12 U.S. investor-owned utilities

	Cost (\$/MWh)	Percentage of total ancillary-service cost
System control	0.18	4
Voltage control	0.51	12
Regulation	0.38	9
Spinning reserve	0.66	16
Supplemental reserve	0.73	18
Energy imbalance	0.47	11
Real-loss replacement	<u>1.22</u>	<u>30</u>
Total	4.15	100

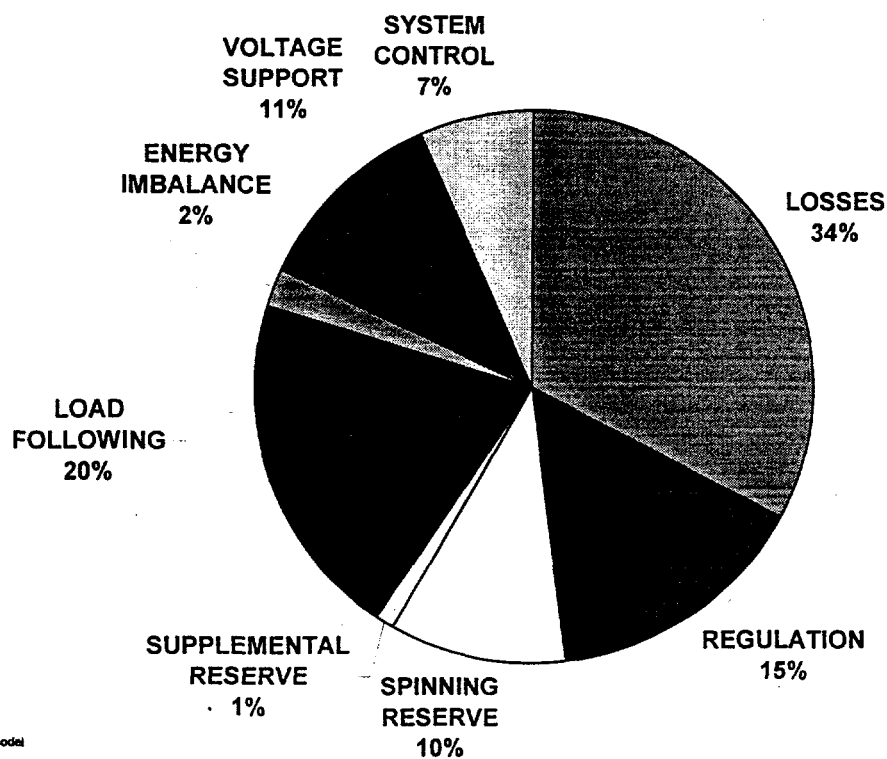


Fig. 1.

The allocation of ancillary-service costs for a hypothetical system. In total, these services cost the equivalent of \$2.3/MWh during the course of a year.

OPERATING-RESERVE BASICS

Electricity is the ultimate “real-time” product, with its production, transportation, and consumption occurring within a fraction of a second. Because electricity moves at nearly the speed of light and cannot readily be stored, bulk-power systems must maintain a continuous and near-instantaneous balance between production and consumption.

Electric utilities have traditionally provided operating reserves as part of the bundled product they sold to their customers. Utilities and their regulators implicitly agreed on the level of reliability to be maintained (and, therefore, on the amount of reserves that each utility carried). In the increasingly competitive electricity industry, transmission-system operators will be commercially independent of generator operation and output sales. Because reliability rules can have substantial effects on the profits from energy production and sales, many entities will closely scrutinize these rules for their effects on their earnings.

This chapter explains how operating reserves are used to quickly restore the electric system to equilibrium after a major outage. It also describes the rules that the North American Electric Reliability Council (NERC) and its regional reliability councils use to specify the minimum levels of operating reserves that each electrical system must maintain, and it presents data on generator outages and system responses for a power pool (New York), a large Midwestern utility (American Electric Power, AEP), and a reliability region (ERCOT). Chapter 6 discusses several issues related to the technical basis for operating-reserve requirements, payment for operating reserves, future data needs, and disaggregation of the several purposes that operating reserves currently serve.

APPLICATION OF OPERATING RESERVES

Figure 2 illustrates how the electric system is intended to operate when a major generating unit suddenly trips offline (Virmani, Lo, and McNair 1992). Prior to the outage, system frequency is very close to its 60-Hz reference value. Generally, within a second after the outage occurs, frequency drops, in this case to just over 59.9 Hz. The frequency decline is arrested primarily because many electrical loads (e.g., motors) are frequency responsive; that is, their demand varies with system frequency. Once the frequency decline exceeds the deadband of the generator governors, the governors at those generators so equipped sense the frequency decline and open valves on the steam turbines, which rapidly increases generator output. After a few more seconds, generator output declines slightly because the higher steam flow through the turbine is not matched by the steam flow from the boiler to the turbine. At this point, the operating reserves, in response to AGC signals from the control center, kick in. More

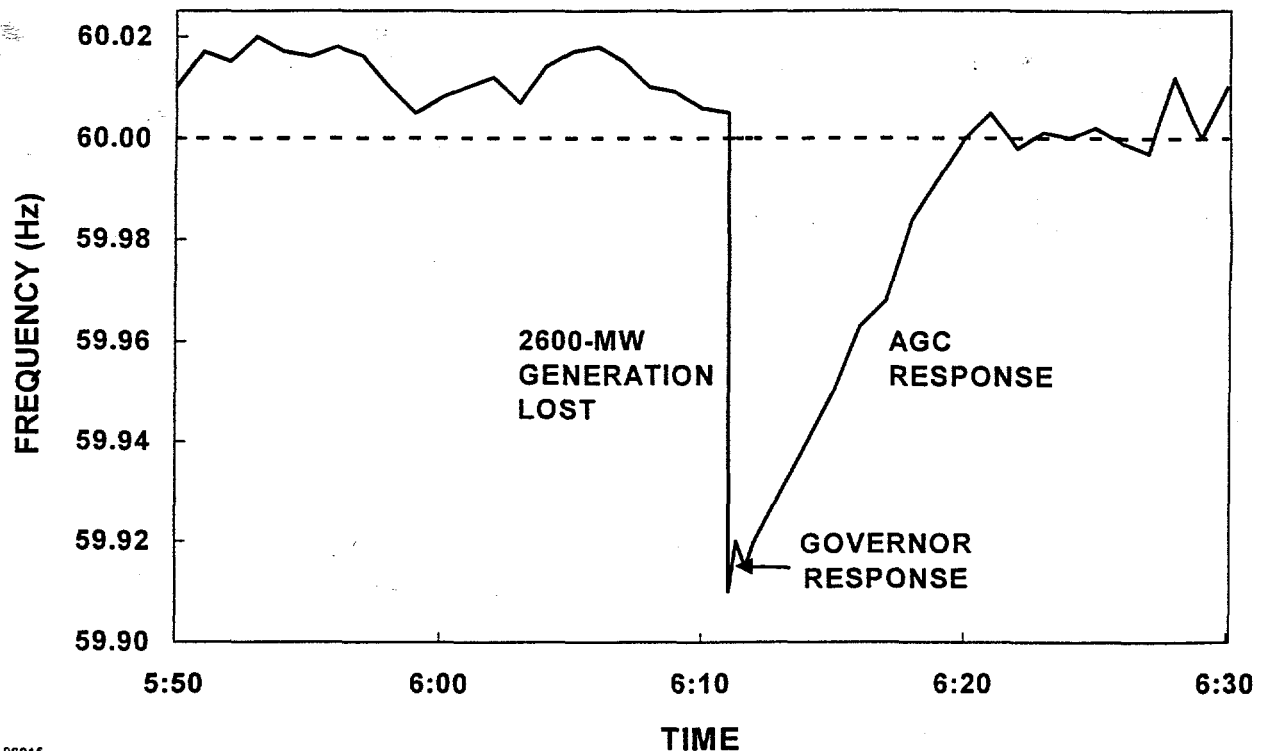


Fig. 2. Interconnection frequency before and after the loss of a 2600-MW generating unit.

fuel is added to the boiler, leading to a higher rate of steam production, which leads to higher power output. In the Figure-2 example, the system worked as it was intended to, and frequency was restored to its pre-contingency 60-Hz reference value within the required ten minutes.

NERC AND REGIONAL DEFINITIONS AND REQUIREMENTS

To understand the purpose and use of operating reserves, one must first understand the concept of a control area and its functions. NERC (1995) defines a control area as: "An electrical system bounded by interconnection (tie line) metering and telemetry. It controls its generation directly to maintain its interchange schedule with other control areas and contributes to frequency regulation of the Interconnection."

Control areas seek to minimize any adverse effect they might have on other control areas within the interconnection by minimizing their ACE. The ACE equation, in slightly simplified form, is:

$$ACE = (I_A - I_S) - 10\beta(F_A - F_S),$$

where I refers to the algebraic sum of all real power (MW) flows on the tielines between an individual control area and its surrounding control areas, F is the interconnection frequency (Hz), A is actual, S is scheduled, and β is the control area's frequency bias (MW/0.1Hz). The

first term shows how well the control area performs in matching its schedules with other control areas (i.e., how well it matches its generation plus net incoming scheduled flows to its loads). The second term is the individual control area's contribution to the interconnection to maintain frequency at its scheduled value (usually 60 Hz). Thus, ACE is the instantaneous difference between actual and scheduled interchange, taking into account the effects of frequency.

Until recently, NERC's (1995) standards for "disturbance conditions" consisted of two elements:

B1 Standard. The ACE must return to zero within ten minutes following the start of the disturbance.

B2 Standard. The ACE must start to return to zero within one minute following the start of the disturbance.

The 10-minute full-response requirement associated with operating reserves derives from the B1 standard, while the need for spinning reserve derives from the B2 standard.

NERC's (1996b) new Disturbance Control Standard replaces the B1 and B2 standards with a single quantitative measure of control-area performance. Control areas must calculate this recovery factor (R) for all disturbances within the range of 80% to 100% of the control area's most severe single contingency. Control areas are expected to meet the new standard 100% of the time for these reportable disturbances. The recovery factor is defined as:

$$R_i = 100\% \times [MW_{LOSS} - \text{Max}(0, ACE_A - ACE_M)] / MW_{LOSS} ,$$

where:

MW_{LOSS}	is the MW size of the disturbance as measured at the beginning of the loss,
ACE_A	is the pre-disturbance ACE if $ACE_A < 0$ and 0 if $ACE_A \geq 0$,
ACE_M	is the maximum algebraic value of ACE measured within the ten minutes following the disturbance event.

The minimum operating-reserve requirements differ from region to region (Interconnected Operations Services Working Group 1997). In the East Central Area Reliability Coordination Agreement (ECAR), the spinning and supplemental reserve requirements are both 3% of the daily peak load. In MAAC, the spinning reserve must be the greater of 700 MW or the capacity of the largest unit on line; its supplemental reserve requirement is 1700 MW. In Florida, the spinning reserve must equal 25% of the largest unit online, and the supplemental reserve must equal 75% of the largest unit. In ERCOT (1997), the responsive-reserve requirement is a set amount, 2300 MW. In the other regions, the requirement is based on either

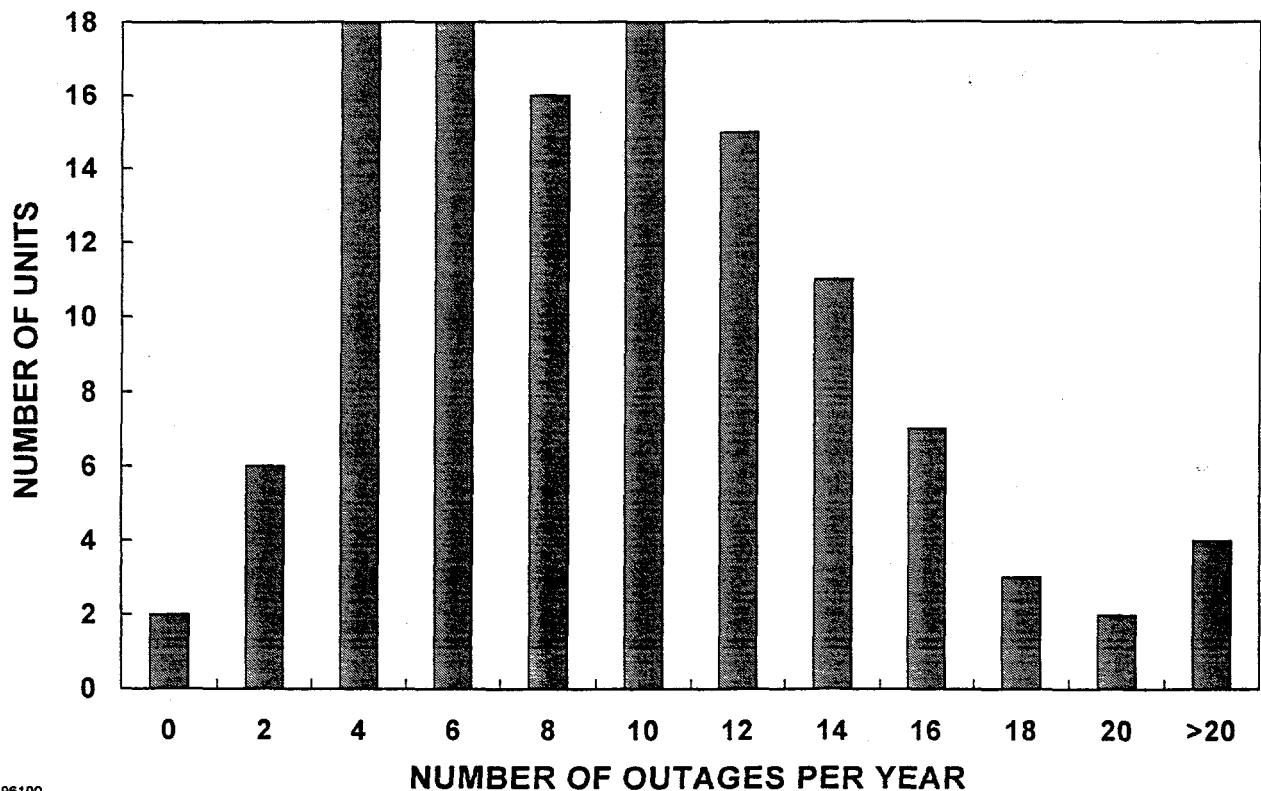
the largest generating unit online or the single most severe contingency (either a large generator or a critical transmission line).

The requirement for minimum levels of operating reserves could be based on either probabilistic or deterministic calculations. In most regions, the requirement is deterministically calculated based on the $N - 1$ criterion. Thus, these requirements are independent of the reliability performance of the generating units in the particular region. The Western Systems Coordinating Council (1997) is the only region with an operating-reserve requirement that recognizes differences among generator types. It requires reserves equal to 5% of the load supplied by hydroelectric resources plus 7% of the load supplied by thermal generation. However, no region incorporates the reliability of *individual* units in its determination of the minimum operating-reserve requirement.

Historically, it may not have been important to recognize the frequency of forced outages at individual generators because utilities provided operating reserves from their portfolio of generating resources and sold this service as part of the bundled electricity product. In the future, with generation unbundled from transmission and with generation increasingly competitive, system operators may need to recognize differences in reliability among generators. These differences could affect the amount of reserves required, how reserves are deployed, and the allocation of costs among generators. Figure 3 shows the number of forced outages for coal-fired units between 600 and 800 MW for 1996 (Curley 1997). Although, on average, these 120 units experienced nine outages a year, two units had none, while at the other extreme, five units had 20 or more outages.

In many cases, the regional requirements for operating reserves include services beyond those associated with protection against generator and transmission outages. Specifically, these reserves are typically used to account for load-forecasting error and sometimes include the generation assigned to regulation. For example, ERCOT (1997) defines operating reserve as "that capability above system demand required to provide for regulation, load forecasting error, and equipment forced outages."

Finally, only limited technical support for the minimum operating-reserve requirements set by the various regions is publicly available. ERCOT (1987) conducted a study ten years ago that demonstrated the adequacy of its spinning-reserve requirement. The study analyzed the stability, voltages, power flows, and frequency at various locations in Texas for the first 15 seconds after an outage occurs. The PJM Interconnection provided several studies, the most recent of which was completed 22 years ago. System operators frequently conduct such contingency studies, but apparently these studies are rarely published.



96100

Fig. 3. Number of forced outages in 1996 for 120 coal-fired generating units between 600 and 800 MW in size.

DATA ON GENERATOR OUTAGES

Several utilities, power pools, and regional reliability councils were contacted to obtain data on outages and the time to recover from those outages. Specifically, data for each control area on the following variables at the time of each outage was requested:

- the number of generating units online
- the number of units providing spinning and supplemental reserves
- the amount of generating capacity committed to operating reserves
- the amount of generation lost because of the outage
- the amount of operating reserves lost because of the outage
- the current hourly and daily peak loads
- the current spot price of electricity (or system lambda)
- the time to restore ACE and frequency to their pre-disturbance values

Ultimately, data were obtained for only three control areas. Discussions with system operators and planners uncovered several reasons why such data often could not be provided. In many cases, these people were extremely busy. As examples, they were developing or responding to new NERC and regional reliability council requirements for system operation and control, they were developing or responding to proposals to create ISOs, or they were

implementing electronic-information systems. As one respondent wrote, "we do not presently have sufficient staff resources to devote toward supplying the data you are requesting as much of it is in a form that would require significant effort to compile."

In other cases, the data were considered proprietary and not for public release. In every case, many of the data elements requested were simply not available (although most of the people contacted agreed that the list shown above is a reasonable one). In several cases, even the most basic data were not made available. As one utility staffer put it "We have two problems here at ... , the accuracy of the data and our inability to store it for long periods of time. We are working on these issues."

Table 6 summarizes the data obtained from NYPP, ERCOT, and AEP; see Hirst and Kirby (1997c) for additional detail on these data sets and the analyses thereof. Normalized by system peak load, AEP has more than four times as many large (greater than 500 MW) outages as does ERCOT and about twice as many outages as does NYPP.

Table 6. Size and number of generator outages (> 500 MW) from three systems

	NYPP	ERCOT ^a	AEP
Peak load (MW)	27,000	48,000	20,000
1995	20	Not available	38
1996	15	21	39

^aDuring the first half of 1997, ERCOT had 10 generator outages greater than 500 MW.

These results show substantial differences in the number of, extent of, and recovery from generator outages (Fig. 4). Relative to the minimum operating-reserve requirement, the outages in ERCOT and New York are much smaller than those for AEP.

AEP not only had larger outages, it also had much longer recovery times. For example, ERCOT averaged 8 outages per year for which the recovery time exceeded the 10-minute standard, NYPP averaged less than one such outage, while AEP averaged 32 such outages. AEP's longer recovery times may be related to the fact that a large percentage of its generation comes from big units; specifically 66% of its system capacity is from units with capacities of 500 MW or more, compared with 50% for ERCOT and 26% for NYPP.

Analysis of these data show that the size of an outage is a statistically significant determinant of outage-recovery time (Fig. 4). [Bilke (1997) analyzed 167 outages in the Eastern and ERCOT interconnections and also found a statistically significant relationship between outage size and recovery time; recovery time increases by 0.4 minutes for every 100-MW increase in outage size.] These data also show, as does Bilke's analysis, that many factors beyond outage size affect recovery times. These other factors likely relate to the system

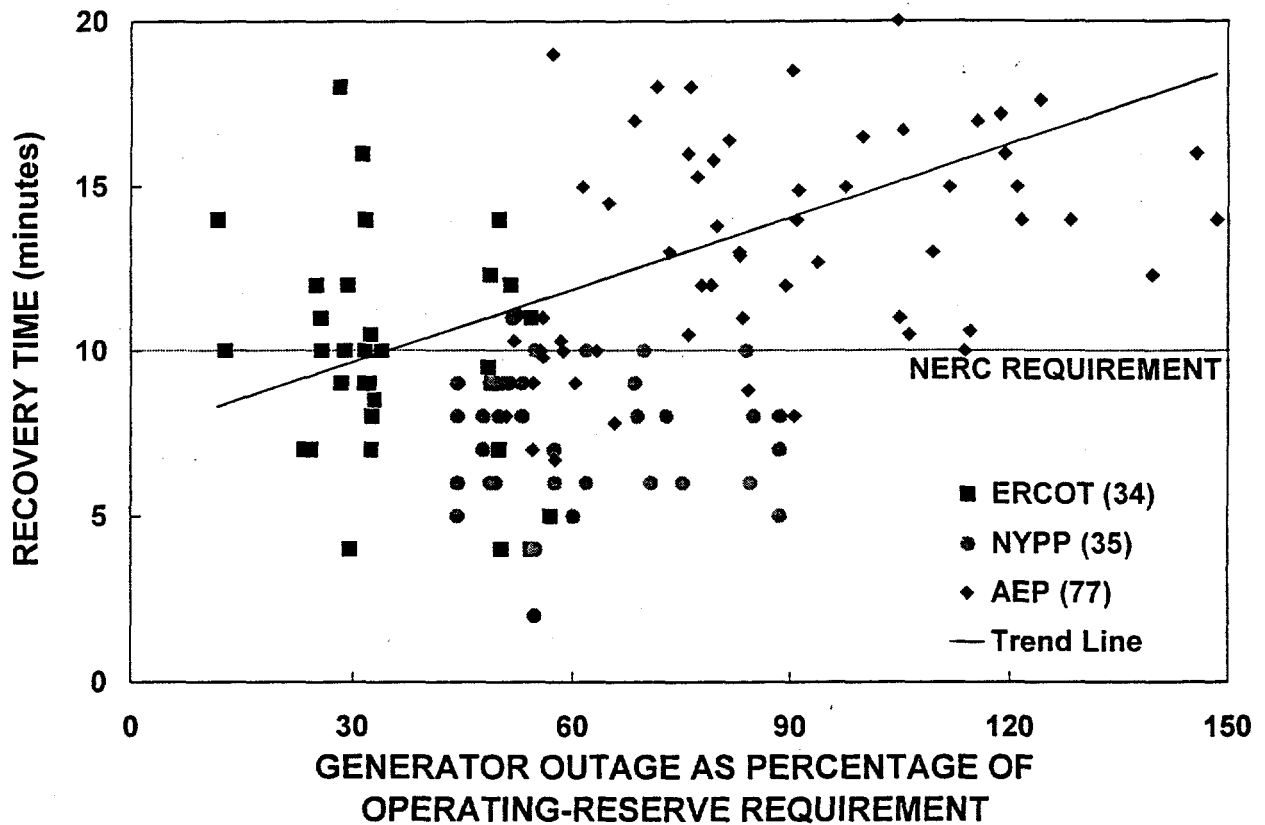


Fig. 4. Time to recover from a generator outage as a function of the outage magnitude relative to the minimum reserve requirement for three systems.

conditions at the time of the outage, including the factors listed earlier in this section. Finally, these data, as well as those analyzed by Bilke, show that many outages exceed the NERC 10-minute requirement and will therefore lead to less than 100% compliance with the new NERC disturbance control standard.

This discussion suggests that the costs and benefits of different amounts of operating reserves are not well documented and may not be well understood. Decisions on the maximum allowable recovery time and the amount of operating reserves to maintain should be examined anew for their applicability to competitive electricity markets.

EMERGING ISSUES ON OPERATING RESERVES

TECHNICAL BASIS FOR RESERVE REQUIREMENTS

NERC and the regional reliability councils are in the midst of a major transformation from a system of voluntary compliance to one of mandatory compliance with reliability requirements. The new Disturbance Control Standard is an important step in the direction of developing and implementing measurable reliability-related performance standards.

Unfortunately, there appears to be only a limited analytical basis for either the 10-minute recovery-period requirement or for the minimum-operating-reserve requirements specified by the ten reliability councils. Although discussions were held with many people at NERC, the councils, power pools, and individual utilities, only two sets of studies were made available. And these studies were 10 to 40 years old. These requirements are based primarily on the experience and judgment of control-area operators and electrical engineers rather than on data and analysis. For example, there appears to be no basis for the requirement that control areas calculate the disturbance control standard only for those outages that fall between 80 and 100% of the most severe single contingency. Limited analysis of data from three control areas shows that the recovery times depend strongly on outage size as well as other factors (Fig. 4).

Although such rules of thumb were sufficient in the past, they may not be enough in the future. In the future, customers will be required to pay explicitly for these operating reserves (because their costs will no longer be buried within the bundled electric rate they formerly paid). Customers, especially large industrial customers, and power marketers will want clear evidence that they are getting services commensurate with what they are paying for those operating-reserve services. And generators that impose less of an operating-reserve burden on the system will want to be compensated accordingly.

Operating-reserve requirements could be based on either probabilistic or deterministic calculations. Ideally, the reliability councils would conduct both types of analyses, testing the results against different sensitivities related to unit-specific forced-outage rates, capacity costs, transmission constraints, and other factors. In general, these requirements are currently based primarily on the magnitude of the largest single contingency. The thinking behind this approach is that the system must be able to withstand such a contingency regardless of the probability of its occurrence. In other words, even if the largest generator has an excellent reliability record (e.g., less than one forced outage a year), the consequences of such an outage are so severe that the system must be protected against its occurrence.

An alternative approach, which merits additional consideration, would focus on the statistics of outages, their severity, and their consequences. In such a probabilistic approach, the performance of individual generators would figure prominently in the determination of the minimum amount, type (capital vs operating costs), and location of operating reserves required.

PAYING FOR RESERVES

The current systems for assigning operating reserves do not distinguish among the performance of *individual* generators. In the future, the system may change in two ways.

- The operating-reserve requirements may increasingly be based on probabilistic in addition to today's deterministic methods. In several regions, the reserve requirement is based on the capacity of the largest unit online. This requirement, however, is completely independent of the reliability of that unit. Perhaps, the amount and type of reserves required at any given time should depend on which generators are online. If that mix of generators is highly reliable, the operating-reserve requirements will be modest. On the other hand, if some of those generators have frequent forced outages, the overall operating-reserve requirement may be higher and may consist of a different mix of generators.
- Although customers will continue to pay for operating reserves, in the first instance suppliers may pay. Specifically, each supplier could be assigned an amount of operating reserves to provide or obtain as a function of its individual unit performance. And when a particular unit trips offline, that unit could be responsible for any extra payments to generators that provided energy during the 30- or 60-minute period that such reserves operate.

In a competitive market, generating units that are highly reliable (e.g., those on the left side of Fig. 3) might be required to provide or pay for different or less operating reserves than would the units that experience frequent outages. This economic signal would provide the appropriate incentive to generation owners, encouraging them to undertake the amount of maintenance that would just balance the higher cost of providing more and more-expensive operating reserves. In addition, when an outage occurs and operating reserves are called upon, the generator responsible for the outage would pay the extra costs of the units that responded to the outage (i.e., the incremental fuel plus operating costs beyond those associated with the spot-market price for that hour). This pricing approach would eliminate subsidies among generators and would provide further incentives to generator owners to maintain high availability levels at their units.

FUTURE DATA NEEDS

It was surprisingly difficult to obtain even the most basic data on forced outages. Utilities, power pools, and regional reliability councils are often unable to readily provide

information on the frequency and consequences of outages. Accurate and complete data are especially difficult to obtain. Control centers often cannot provide data on the number of units online at the time of the outage, the amount of operating reserves online, and whether the unit that tripped offline was also providing operating reserves. Because a competitive electricity market will likely insist on technically defensible reserve requirements, NERC should collect more data along the following lines: unit output (MW) immediately before the outage, unit capacity (MW), system load at the time of the outage (MW), system operating reserves at the time of the outage (MW), number of units online at the time of the outage, and the time to return ACE and frequency to their precontingency values.

MIXING FUNCTIONS WITHIN OPERATING RESERVES

Currently, operating reserves are often used to protect against load-forecast errors as well as supply outages. It may be inappropriate to assign operating reserves to two disparate functions even when the services are provided by the same generating units.* The load-forecast errors and generator forced outages represent very different phenomena, with the former being a commercial function and the latter a bulk-power-reliability function. The cost of maintaining reserves to protect against generator outages should, as discussed above, fall on those generators responsible for the operating-reserve requirement (e.g., with frequent forced outages or large unit sizes). The cost of maintaining reserves to protect against load-forecast errors should be assigned to those customers or scheduling coordinators that forecast poorly; these costs should appear in the load-following or energy-imbalance service, not in operating reserves.

Although FERC defined separate operating reserve and regulation services, operating reserves often include regulation (the use of generating units to follow minute-to-minute fluctuations in system load). Here, too, these services should be treated separately. Generators that provide regulation grant the system operator the right to vary the outputs of their units up or down, from minute to minute, within a specified range ($\pm x$ MW with a ramp rate of y MW/minute). Generators that provide operating reserves grant the system operator the right to increase (but not decrease) the output from their units when a major disturbance occurs. Based on the data analyzed here, a unit providing operating reserves can expect to be called upon to provide those reserves once or twice a month. Thus, the operation of generators for regulation differs substantially from the operation of those units for operating reserves. Combining these two disparate functions is inappropriate.

*Even though these services (operating reserves, load following, and regulation) may be provided by the same generating units, their requirements, costs, prices, and billing to customers should be considered separately.

VOLTAGE-CONTROL BASICS

Generators and various types of transmission equipment are used to maintain voltages throughout the transmission system. Injecting reactive power into the system raises voltages, and absorbing reactive power lowers voltages. Voltage-support requirements are a function of the location and magnitudes of generator outputs and customer loads. These requirements can differ substantially from location to location and can change rapidly as the locations and magnitudes of generation and load change or if the transmission-system configuration changes. At very low levels of system load, transmission lines act as capacitors and increase voltages. At high loads, however, transmission lines absorb reactive power and lower voltages. Most transmission-system equipment (e.g., capacitors, inductors, and tap-changing transformers) is static and can respond to changes in voltage requirements only slowly and in discrete steps. Some transmission-system equipment (e.g., SVCs) and generators can respond within cycles to changing reactive-power requirements.

Voltages must be controlled for three reasons. First, both customer and power-system equipment is designed to operate within a range of voltages, usually within $\pm 5\%$ of the nominal voltage. At low voltages, many types of equipment perform poorly; light bulbs provide less illumination, induction motors can overheat and be damaged, and some electronic equipment will not operate at all. High voltages can damage equipment and shorten their lifetimes.

Second, reactive power consumes transmission and generation capacity. To maximize the amount of real power that can be transferred across a congested transmission interface, reactive-power flows must be minimized. Similarly, reactive-power production can limit a generator's real-power capability.

Third, moving reactive power on the transmission system incurs real-power losses. Both capacity and energy must be supplied to replace these losses.

Voltage control is complicated by two additional factors. First, the transmission system itself is a nonlinear consumer of reactive power, depending on system loading. System reactive requirements vary in time as load levels and load and generation patterns change.

Second, the bulk-power system is composed of many pieces of equipment, any one of which can fail at any time. Therefore, the system is designed to withstand the loss of any single piece of equipment and to continue operating without impacting any customers. That is, the system is designed to withstand a single contingency.

These two factors result in a dynamic reactive-power requirement. The loss of a generator or a major transmission line can have the compounding effect of reducing the reactive supply and, at the same time, reconfiguring flows such that the system is consuming additional reactive power. At least a portion of the reactive supply must be capable of responding quickly to changing reactive-power demands and to maintain acceptable voltages throughout the system. Thus, just as an electrical system requires real-power reserves to respond to contingencies, so too it must maintain reactive-power reserves.

Loads can also be both real and reactive. The reactive portion of the load could be served from the transmission system. As explained in the Appendix of Kirby and Hirst (1997), reactive loads incur more voltage drop and reactive losses in the transmission system than do similar size (MVA) real loads. Vertically integrated utilities often included charges for provision of reactive power to loads in their rates. With restructuring, the trend is to restrict loads to operation at near zero reactive-power demand (i.e., a power factor* of 1). The California ISO limits loads to power factors between 0.97 lagging (absorbing reactive power) and 0.99 leading (generating reactive power) (Pacific Gas and Electric et al. 1997).

The significant differences between the real and reactive ancillary services are:

- Real power can be delivered over much greater distances so the supplying resources are not as location constrained, whereas reactive resources must be distributed throughout the power system.
- Generation of real power requires the conversion from an energy resource, such as chemical or nuclear fuel, sunlight, or a mechanical resource like wind or water flow, whereas reactive power requires almost no "fuel" to produce.

Much of the complication associated with voltage control comes from combining these various aspects into a single service. Rather than splitting this service along functional lines, FERC and the industry have elected to split the service based on commercial ownership of the resources that supply the service. The requirements and compensation for generation-based resources fall under what FERC calls Voltage Regulation and Reactive Power Management from Generation. Transmission-based resources are addressed under the general transmission tariff. The range of physical requirements (speed of response, need for contingency reserve, etc.) still exist, but they are not addressed in the basic definitions as they are for real-power services.

As with most ancillary services, the need for voltage control stems from an overall system requirement, requires resources capable of supplying that need, and must have a central-control function directing those resources to meet the requirement. Suppliers of the resources

*The power factor (PF) is the ratio of real power (in MW) to apparent power (MVA). Apparent power is related to real and reactive power (MVAR) according to $MVA = \sqrt{[(MW)^2 + (MVAR)^2]}$.

are not able to independently determine the system's voltage-control needs. Only the system operator has sufficient information to know the system requirements, both current and contingency, and to deploy those resources effectively. Similarly, while service requirements result from customer choices in terms of load patterns and generation choices, the customers do not have sufficient information about the configuration of the transmission system or the actions of other customers to know what reactive-power requirements will result ahead of time from their choices. Again, the system operator is needed to deploy resources to meet requirements and to transmit appropriate price signals to customers.

TRANSMISSION-SYSTEM VOLTAGE CONTROL

In moving power from generators to loads, the transmission network introduces both real and reactive losses. Real-power losses arise because aluminum and copper (the materials most often used for transmission lines) are not perfect conductors; they have resistance. The reactive nature of transmission lines is associated with the geometry of the conductors themselves (primarily the radius of the conductor) and the conductor configuration (the distances between each conductor and ground and the distances among conductors).

The reactive-power behavior of transmission lines is complicated by their inductive and capacitive characteristics. As shown in Fig. 5, at low line loadings, the capacitive effect dominates, and generators and transmission-related reactive equipment must absorb reactive power to maintain line voltages within their appropriate limits. On the other hand, at high line loadings, the inductive effect dominates, and generators, capacitors, and other reactive devices must produce reactive power. The balance point at which the inductive and capacitive effects cancel each other (what is called surge-impedance loading) is typically about 40% of the line's thermal capacity.

Figure 5 also shows that at both low and high line loadings (but not around the surge-impedance loading), reactive losses are greater than real losses. At full line loading, reactive losses are five times greater than real losses for a 230-kV line and nine times higher for a 345-kV line. (At 50% of line loading, the factors are two and four for the 230- and 345-kV lines, respectively.)

If uncompensated, these line losses reduce the amount of real power that can be transmitted from generators to loads. Figure 6 shows how transmission-line capacity decreases as the line length increases if there is no voltage support (injection or absorption of reactive power) on the line. At short distances, the line's capacity is limited by thermal considerations; at intermediate distances the limits are related to voltage drop; and beyond roughly 300 to 350 miles, stability limits dominate. An ECAR (1997) study showed that the addition of series capacitors to boost voltages would increase the capacity of a particular transmission line by about 60%.

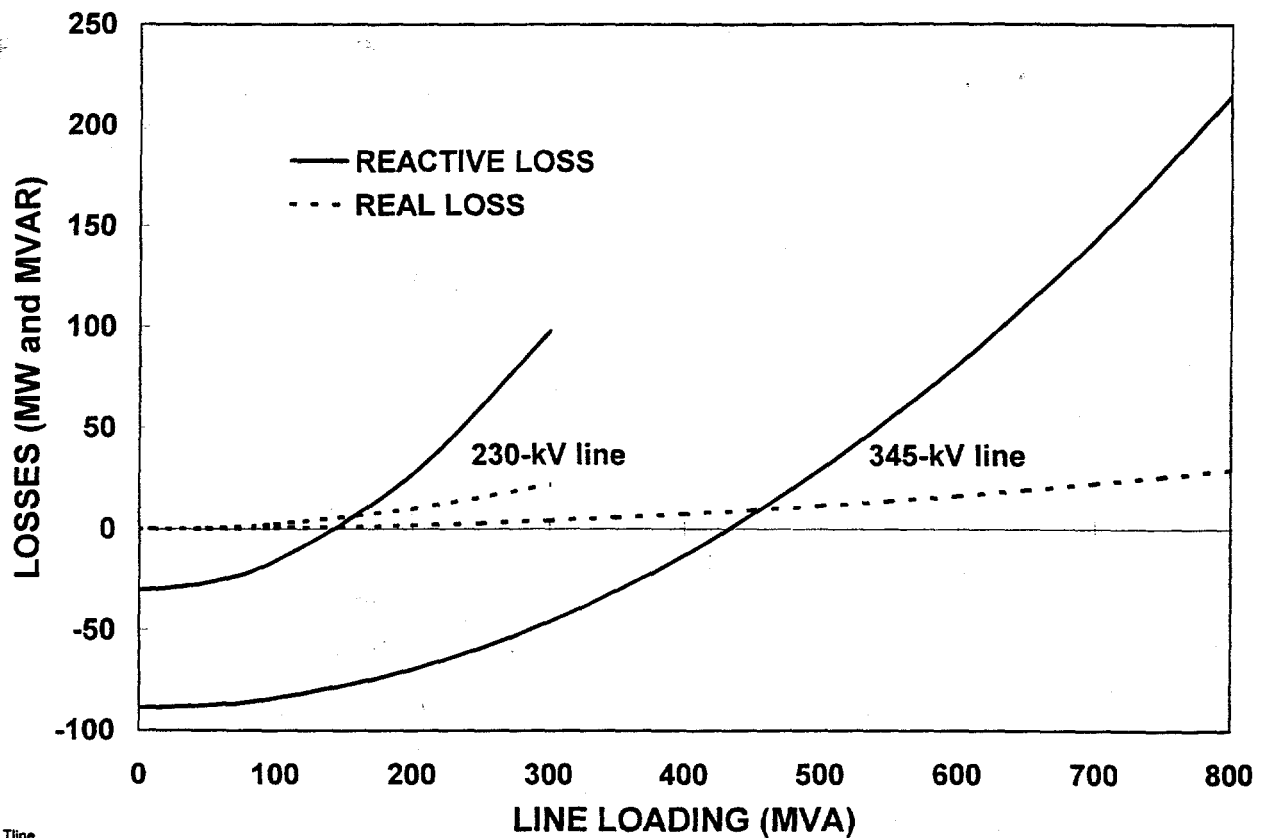


Fig. 5. Transmission lines supply reactive power to the system when lightly loaded but absorb reactive power when heavily loaded. These results are for a 100-mile line with voltages supported at both ends.

VOLTAGE-CONTROL EQUIPMENT

Power-system designers and operators can use various devices to control voltages by injecting, absorbing, or forcing the flow of reactive power. These devices differ in several important characteristics: response speed, continuity of control, response to system voltage changes, and capital and operating costs (Table 7).

Generation: A generator's primary function is to convert fuel (or some other energy resource) into electric power. Almost all synchronous generators also have considerable control over their terminal voltage and reactive-power output.

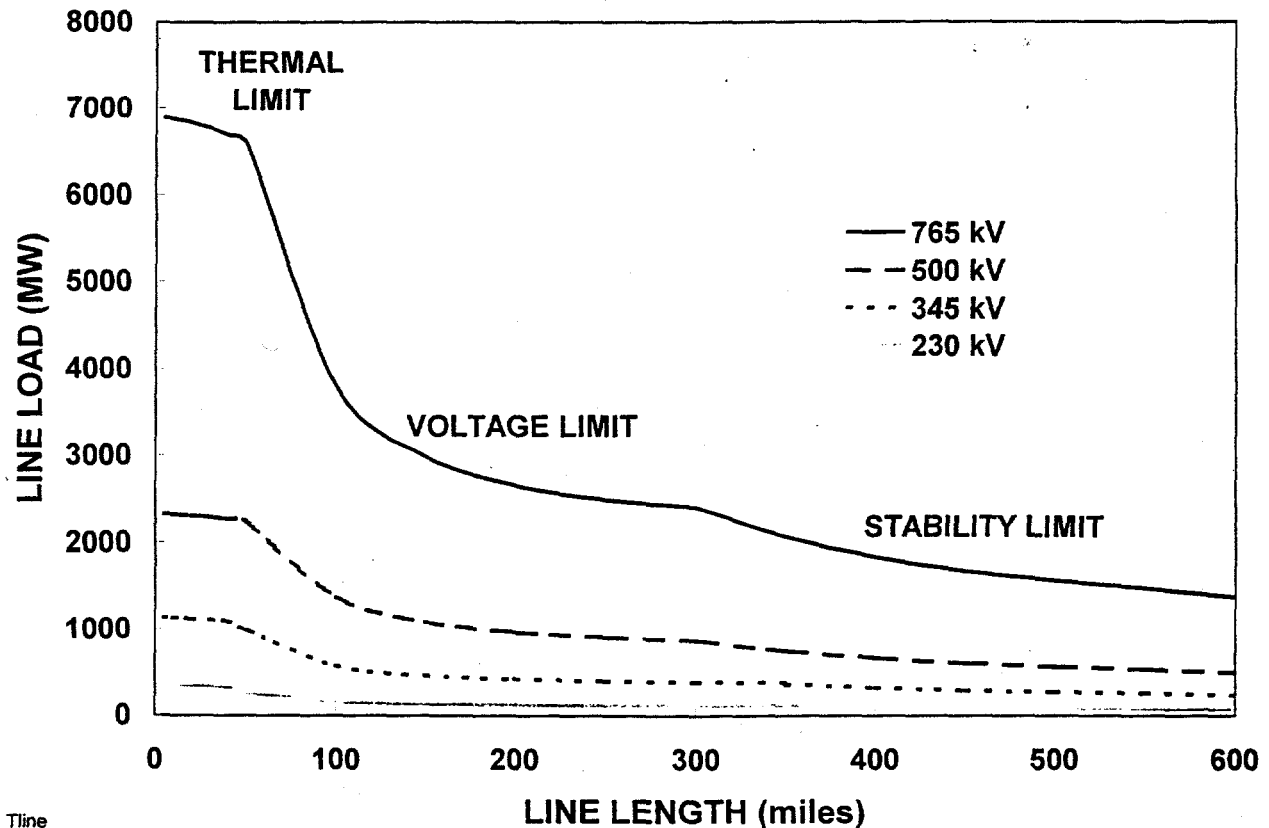


Fig. 6. Absent reactive compensation to maintain voltages, transmission-line capacity is limited to shorter lengths.

The ability of a generator to provide reactive support depends on its real-power production. Figure 7 shows the combined limits on real and reactive production for a typical generator. The generator prime mover (e.g., the steam turbine) is usually designed with less capacity than the electric generator, resulting in the prime-mover limit in Fig. 7. The designers recognize that the generator will be producing reactive power most of the time. This generator, at maximum real-power output, can produce 0.62 MVAR for every megawatt of real-power output, based on the 0.85 lagging power-factor limit. Or it can absorb 0.41 MVAR for every megawatt of real-power output, based on the 0.93 leading power-factor limit.* To produce or absorb additional VARs beyond these limits would require a reduction in the real-power output of the unit.

*The leading power factor limit is 0.93, rather than the 0.95 shown in Fig. 7, because the former is the prime-mover limit. The 0.95 power factor shows the intersection of the core-end heating limit and the armature-heating limit, not the limit on real-power production.

Table 7. Characteristics of voltage-control equipment

Equipment type	Speed of response	Ability to support voltage	Costs		
			Capital (per kVAR)	Operating	Opportunity
Generator	Fast	Excellent, additional short-term capacity	Difficult to separate	High	Yes
Synchronous condenser	Fast	Excellent, additional short-term capacity	\$30-35	High	No
Capacitor	Slow, stepped	Poor, drops with V^2	\$8-10	None	No
SVC	Fast	Poor, drops with V^2	\$45-50	Moderate	No
STATCOM	Fast	Fair, drops with V	\$50-55	Moderate	No
Distributed generation	Fast	Fair, drops with V	Difficult to separate	High	Yes

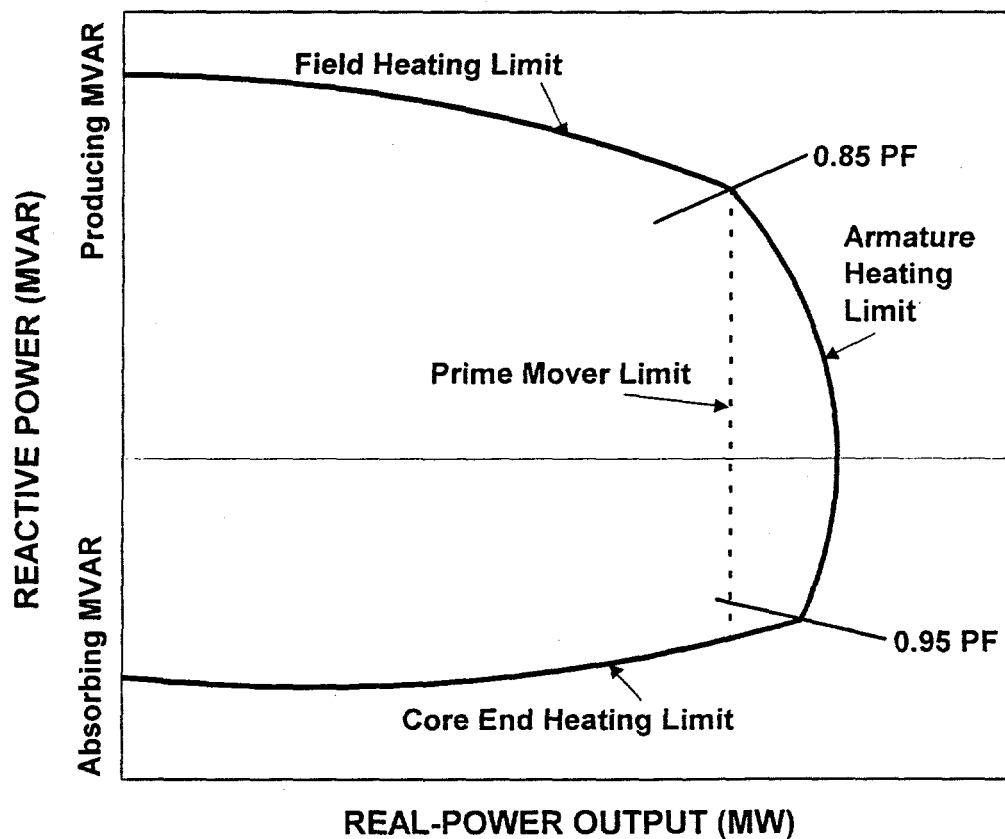


Fig. 7. Reactive-power capability depends on real-power production for a synchronous generator.

Historically, only the capital and operating costs that could be associated with the extra equipment (e.g., parts of the stator, rotor, voltage regulator, and exciter plus the operating costs associated with field losses) required for voltage control were charged to the voltage-control function. In a restructured industry, the opportunity costs associated with reduced real-power sales when excessive reactive power is required will be a key part of the cost of providing voltage control from generation.* Generator reactive-power capability is determined during the design of the machine. Because capital costs are involved in providing this capability, builders of new generating units will likely not include voltage-support capability unless (1) the market or a regulated tariff compensates the generators or (2) voltage-support capability is a requirement of connecting to the grid.

Synchronous Condensers: Every synchronous machine (motor or generator) has the reactive capabilities discussed above. Synchronous motors are occasionally used to provide voltage support to the power system as they provide mechanical power to their load. Some combustion turbines and hydro units are designed to allow the generator to operate without its mechanical power source simply to provide the reactive-power capability to the power system when the real-power generation is unavailable or not needed. Synchronous machines that are designed exclusively to provide reactive support are called synchronous condensers. Synchronous condensers have all of the response speed and controllability advantages of generators without the need to construct the rest of the power plant (e.g., fuel-handling equipment and boilers). Because they are rotating machines with moving parts and auxiliary systems, they may require significantly more maintenance than static alternatives. They also consume real power equal to about 3% of the machine's reactive-power rating. That is, a 50-MVAR synchronous condenser requires about 1.5 MW of real power.

Capacitors and Inductors: Capacitors and inductors are passive devices that generate or absorb reactive power. They accomplish this without real-power losses or significant operating expense. The output of capacitors and inductors is proportional to the square of the voltage. Thus, a capacitor bank (or inductor) rated at 100 MVAR will produce (or absorb) only 90 MVAR when the voltage dips to 95% of its nominal value but it will produce (or absorb) 110 MVAR when the voltage rises by 5%. This relationship is helpful when inductors are employed to hold voltages down. The inductor absorbs more reactive power when voltages are highest and the device is needed most. The relationship is unfortunate for the more common case where capacitors are employed to help support voltages. In the extreme case, voltages fall, and capacitors contribute less, resulting in a further degradation in voltage and even less support from the capacitors; ultimately, voltage collapses and outages occur.

*A recent study suggests that the costs and prices for voltage support will be highly nonlinear with system load (Hirst and Kirby 1997b). At very high loads, the opportunity cost of voltage support will far exceed the embedded cost because of the required cut in real-power output and the likely high spot energy price during such peak-demand periods.

Inductors are discrete devices designed to absorb a specific amount of reactive power at a specific voltage. They can be switched on or off but offer no variable control.

Capacitor banks are composed of individual capacitor cans. The cans are connected in series and parallel to obtain the desired capacitor-bank voltage and capacity rating. Like inductors, capacitor banks are discrete devices but they are often configured with several steps to provide a limited amount of control.

Static VAR Compensators (SVCs): An SVC combines conventional capacitors and inductors with fast switching capability. Switching takes place in the subcycle timeframe (i.e., in less than 1/60 of a second), providing a continuous range of control. The range can be designed to span from absorbing to generating reactive power. Consequently, the controls can be designed to provide very fast and effective reactive support.

Because SVCs are based upon capacitors, they suffer from the same degradation in reactive capability as voltage drops. They also do not have the short-term overload capability of generators and synchronous condensers. SVC applications usually require harmonic filters to reduce the amount of harmonics injected into the power system.

Static Synchronous Compensators (STATCOMs): The STATCOM is a solid-state shunt device that generates or absorbs reactive power and is one member the FACTS family (Kessinger 1997). The STATCOM is similar to the SVC in response speed, control capabilities, and the use of power electronics. Rather than using conventional capacitors and inductors combined with fast switches, however, the STATCOM uses power electronics to synthesize the reactive output. Consequently, output capability is generally symmetric, providing as much capability for production as absorption. The solid-state nature of the STATCOM means that, similar to the SVC, the controls can be designed to provide very fast and effective voltage control.

While not having the short-term overload capability of generators and synchronous condensers, STATCOM capacity does not suffer as seriously as SVCs and capacitors do from degraded voltage. STATCOMs are current limited so their MVAR capability responds linearly to voltage as opposed to the voltage-squared relationship of SVCs and capacitors. This attribute greatly increases the usefulness of STATCOMs in preventing voltage collapse.

Distributed Generation: Distributing generation resources throughout the power system can yield benefits if the generators have the ability to supply reactive power. Without this reactive-power ability, performance of the transmission and distribution system can be degraded.

Induction generators are an attractive choice for small grid-connected generation, primarily because they are relatively inexpensive, are easy to synchronize, and have mechanical characteristics useful for some applications (e.g., wind). However, they absorb reactive power, cannot generate reactive power, and are not controllable. If the generator output fluctuates (as wind does), the reactive demand of the generator fluctuates as well, compounding voltage-

control problems for the transmission system. Induction generators can be compensated for with static capacitors, but this does not address the fluctuation problem or provide controlled voltage support.

Many distributed generation resources are now being coupled to the grid through solid-state power electronics, to allow the prime mover's speed to vary independently of the power system frequency. Photovoltaics generate direct current and require inverters to couple them to the power system. Energy-storage devices (e.g., batteries, flywheels, and superconducting magnetic-storage devices) require solid-state inverters to interface with the grid. This increased use of a solid-state interface between the device and the power system offers full reactive-power control, similar to that of a STATCOM. In fact, most devices do not have to be providing real power for the full range of reactive power to be available. The generation resource can be out of service while the reactive component is fully functional. Development of solid-state power electronics has turned a potential problem into a benefit, allowing distributed resources to contribute to voltage control.

Transformers: Transformers provide the capability to raise alternating-current generation voltages to levels that make long-distance power transfers practical and then lower voltages back to levels that can be distributed and used. The primary coils of the transformer convert electric power into a magnetic field circulating in an iron core. The secondary coils reconvert the magnetic field into electric power. The ratio of the number of turns in the primary to the number of turns in the secondary determines the ratio of the primary voltage to the secondary voltage. By tapping the coils at various points, the ratio between the primary and secondary voltages can be adjusted. Transformer taps can be either fixed or adjustable under load through the use of a load-tap changer. Tap capability is selected for each application during transformer design. Fixed or variable taps often provide $\pm 10\%$ voltage selection with fixed taps in 5 steps and variable taps in 32 steps.

Transformer-tap changers can be used for voltage control, but the control differs from that provided by reactive-power sources. Transformer taps can force voltage up (or down) on one side of a transformer, but only at the expense of reducing (or raising) voltage on the other side. The reactive-power required to raise (or lower) voltage on a bus is forced to flow through the transformer from the bus on the other side.

The transformer must be taken out of service and de-energized to adjust the fixed taps, a time consuming and inconvenient process. Consequently, fixed taps are useful when compensating for load growth and other long-term changes in system use. Load-tap changers are used for more rapid adjustments, such as compensating for the voltage fluctuations associated with the daily load cycle.

Differences among Equipment Types: Generators, synchronous condensers, SVCs, and STATCOMs all provide fast, continuously controllable voltage control. Load-tap changer transformers provide nearly continuous voltage control but they are slow. Because the

transformer moves reactive power from one bus to another, the control gained at one bus is at the expense of the other. Capacitors and inductors are not variable and offer control only in large steps.

An unfortunate characteristic of capacitors and capacitor-based SVCs is that output drops dramatically when voltage is low and support is needed most. Their capability is proportional to the square of the terminal voltage. STATCOMs provide more support under low-voltage conditions than do capacitors or SVCs because their capability drops linearly with voltage. The output of rotating machinery (i.e., generators and synchronous condensers) generally rises with dropping voltage. Generators and synchronous condensers generally have additional emergency capacity that can be used for a limited time.

Voltage-control characteristics favor the use of generators and synchronous condensers. Costs, on the other hand, favor capacitors. Generators have extremely high capital costs because they are designed to produce real power, not reactive power. Even the incremental cost of reactive support from generators is high, although it is difficult to unambiguously separate reactive-power costs from real-power costs. Operating costs for generators are high as well because they can involve real-power losses. Finally, because generators have other uses, they experience opportunity costs when called on to provide simultaneously high levels of reactive and real power. Synchronous condensers have the same costs as generators; but, because they are built solely to provide reactive support, their capital costs are not as high and they incur no opportunity costs. SVCs and STATCOMs are high-cost devices as well, although their operating costs are lower than those for synchronous condensers and generators.

EMERGING ISSUES ON VOLTAGE CONTROL

System operation has three objectives when managing reactive power and voltages. First, it must maintain adequate voltages throughout the transmission system for both current and contingency conditions. Second, it seeks to minimize congestion of real-power flows. Third, it seeks to minimize real-power losses. These were the system-control objectives before restructuring, and they will continue to be the objectives after restructuring of the U.S. electricity industry.

Restructuring will almost surely change the mechanisms that system operators use to acquire and deploy reactive-power resources. Historically, the system operator, generators, transmission devices, and transmission system were all owned and operated by the same entity. In the future, the entities that own and operate generation may differ from those that own transmission, and both may differ from the system operator (which likely will control the transmission system). These changes will require the creation of market structures plus requirements for connection to the transmission system and the operation of certain equipment. In spite of these changes in requirements, the system operator must have the authority to deploy resources to meet the system's voltage-control objectives.

Not surprisingly, more progress has been made in addressing the technical objectives of managing voltages than in developing market structures to perform this function in a restructured industry. NERC (1996a) guidance focuses on the concept of control-area responsibility for controlling voltage and reactive resources, allowing control areas to develop appropriate rules. Control areas are to maintain sufficient reactive resources to support voltages under first-contingency conditions. Each control area should take care of its own needs and avoid placing a burden on other control areas. NERC policies offer little guidance on what resources should be available or how the system operator should acquire and deploy them, leaving these decisions to the regional reliability councils and the individual control areas.

TECHNICAL GUIDANCE AND COMPENSATION RULES

Regional councils differ in the level of detailed guidance they offer their members. The Southwest Power Pool (1997), the Southeastern Electric Reliability Council (1995), and the Florida Reliability Coordinating Council (1996) provide only general guidance to their members on maintaining adequate voltage support during normal and contingency conditions. Individual control-area operators use the authority of these directives to order specific actions directed at voltage support. ERCOT (1997) provides additional direction on required generator

capability and authorizes control-area operators to use each generator's full reactive capability to maintain system voltages.

Regions containing tight power pools, such as the Mid-Atlantic Area Council and the Northeast Power Coordinating Council along with their member pools, have more detailed rules concerning both generator-capability requirements and system operations to maintain voltages (PJM 1997; Northeast Power Coordinating Council 1996; New England Power Pool 1996). These rules govern which conditions to study when determining voltage-support requirements and generator-reactive-capability requirements as well as their use. Rules governing voltage-control actions to be taken by system operators generally use the lowest-cost resources first and delay expensive actions, such as redispatching generator real-power output. When the problem is addressed at all, generators are typically compensated at cost if they are required to adjust their real-power output to support voltages. The Northeast Power Coordinating Council (1992) operating guide provides a typical example of the order of control actions:

Low-cost transmission-system and generation-control actions

- adjust transformer taps
- switch capacitors and reactors
- adjust SVCs
- use full reactive capability of online generators

Higher-cost adjustments to economical energy production

- start additional generation
- modify economy transactions with other areas and/or deviate from economic dispatch
- operate hydro units as synchronous condensers, where possible
- reschedule pumped-hydro units to generate or pump
- purchase energy

Adjustments to the transmission-system configuration

- switch out transmission lines if security limits will not be violated

None of the previously mentioned organizations has fully addressed voltage control in a restructured, competitive generation market. California's efforts at restructuring have extended, or at least made more explicit, the relationships among the various parties when providing reactive power (Pacific Gas and Electric et al. 1997). California's ISO will be charged with the planning and operating responsibilities associated with maintaining reliability and facilitating markets. One of these responsibilities involves assuring the adequacy and reliability of voltage support. But the California ISO has no generation resources of its own (it is independent of all generation). The ISO obtains reactive-power-generation resources by making them a condition of connection to the grid. All generators are required to have a continuous operating range from 0.90 power factor lag to 0.95 power factor lead. The ISO has the right to direct generators—without compensation—to operate anywhere within this range.

Only if the ISO requires a generator to operate outside this range will the ISO compensate the generator.

REQUIREMENTS IN A RESTRUCTURED INDUSTRY

ISO management of reactive resources involves equipment capability and actual delivery. That is, the ISO must acquire (either by mandate or purchase) the right to specified amounts of reactive-power capability. That power must then be delivered (or absorbed) in response to the time-varying and location-specific requirements of the system.

The system operator uses computer models to help determine what voltage-control resources are needed and, to a lesser extent, how to deploy those resources most efficiently. A constrained optimal-power-flow analysis can determine the reactive-power dispatch that minimizes real-power losses or can calculate location-specific reactive-power prices (El-Keib and Ma 1997). Including physical limits, bus voltage requirements, and contingencies as constraining conditions assures that the calculated solutions meet the physical requirements. If all costs are presented to the system operator, the optimal power flow can be used to find a least-cost solution.

Application of alternative objective functions can lead to very different uses of reactive-power devices. According to one analysis, traditional VAR dispatch seeks to minimize control *actions* (Dandachi et al. 1996). Minimizing the *cost* of voltage control could lower voltage-support costs by about one fourth. The resulting solution, however, leads to an entirely different voltage profile that requires many more control actions, such as adjustments of transformer taps.

In a competitive environment, determination of who has provided what service will be more important. Use of claimed capability may not be adequate. The actual VAR output from a generator may not match the manufacturer's capability specifications. Public Service Company of Colorado found that operating limits often prevent generators from providing full rated VAR output. Adjustment of tap settings on the step-up (and other) transformers, adjustment of station service voltage levels, and recalibrating and setting alarms and meters are often needed to increase the actual VAR output. Better training of operators on the use of voltage-control equipment is also important. This utility achieved a 500-MVAR increase in actual capability at 13 generating units, a 50% increase in output (Panvini and Yohn 1995).

Because actual and predicted performance might not match well, ECAR (1996) developed methods for rating the real- and reactive-power performance of generating units under normal operating conditions. The reactive-power tests must be conducted at least once every two years and require maintenance of scheduled voltages for at least two hours. ISOs will likely develop and implement similar methods to be sure that the reactive supplies contracted to the ISO can and do perform as specified. A corresponding set of penalties must also be developed to deter persistent noncompliance.

ISOs can acquire reactive resources either through mandates or purchases. The California ISO (as noted above) chose the mandatory approach. Other ISOs plan to pay generators their embedded costs for reactive resources. To the extent that reactive supplies are not geographically restricted, it might be possible to create competitive markets for the acquisition of such resources. Many believe that the location limitations on reactive resources are sufficiently great that competitive markets cannot develop for this service. Unfortunately, determining the embedded cost of generator-provided reactive power is ambiguous at best. This ambiguity occurs because the same equipment is used to provide both real and reactive power. What percentages, for example, of the exciter, generator stator, generator rotor, turbine assembly, and step-up transformer should be assigned to each function?

A Southern Company proceeding before FERC illustrates well these complexities. Southern proposed to assign some of the cost of the turbines and 100% of the costs of the exciter and its cooling system to reactive power. FERC (1997b) disagreed and assigned 100% of the turbine cost to real-power production. Southern computed its reactive-power charges on the basis of a pair of load-flow studies, a base case and a transaction case. Southern used the difference in reactive-power requirements between the two cases as the basis for its proposed charges. FERC rejected the use of *incremental* pricing for generator reactive power because Southern was using *embedded-cost* pricing for transmission-supplied reactive power.

Finally, the ISOs and the other participants in bulk-power markets will need to decide which entity has the responsibility and authority to plan for and determine the need for (and location of) additional reactive resources and how those resources are to be used. Some utilities rely primarily on transmission equipment to maintain voltages under normal operating conditions; these utilities use their generator reactive capabilities only to respond to contingencies (Nedwick, Mistr, and Croasdale 1995). Other utilities rely on their generators to manage voltages under normal operating conditions. Who, in the future, will make these kinds of decisions?

CONCLUSIONS

Ancillary services, traditionally buried within the operations of vertically integrated utilities, play vital roles in today's bulk-power systems. Services such as regulation and load following maintain the system's generation/load balance and help to prevent outages. Other services, such as operating reserves and system-black-start capability, help to promptly restore service when outages occur. Still other services, including backup supply and dynamic scheduling, serve primarily commercial functions and support the creation of competitive bulk-power markets.

FERC, the Interconnected Operations Services Working Group, Oak Ridge National Laboratory, and various entities developing ISO proposals have made substantial progress in identifying and defining the key ancillary services. However, much work remains to be done to measure the supply and consumption of these services and to design market structures that will facilitate efficient and competitive markets for some of these services.

Unbundling has no value in its own right. It makes sense to unbundle and separately price services only when the benefits exceed the metering, accounting, billing, and other transaction costs of unbundling. Unbundling ancillary services can improve economic efficiency by providing information to both suppliers and consumers. Suppliers will provide only those services that they can deliver profitably. For example, the output at hydro units can be changed very rapidly, which makes them ideal units to provide regulation; combustion turbines can be started and brought to full output within a few minutes, which makes them ideal units to provide supplemental reserve. And consumers will want to purchase only those services for which the benefits exceed the costs. These considerations require that the services be well defined with a minimum of overlap. Unbundling generation and transmission services should lower costs to consumers even though it results in a loss of economies of scope.

Spinning and supplemental operating reserves provide an example of definitional ambiguities that may complicate formation of competitive markets for their provision and use. Operating reserves are typically used both to protect against generator and transmission outages and to protect against load-forecast errors. The load-forecast errors and supply-side outages represent very different phenomena, and the costs of responding to these disparate factors should probably be assigned to those responsible for each factor. Specifically, the cost of maintaining reserves should properly fall on those generators that cause these costs to be incurred (e.g., those with frequent forced outages or large sizes). And the cost of maintaining reserves to protect against load-forecast errors should be assigned to those customers and suppliers that fail to meet their schedules; these costs should show up in the load-following or

energy-imbalance services. A spot market could address poor load forecasting in real time but it would not be able to respond to a generation or transmission failure in real time.

Markets will likely play a growing role in setting prices for ancillary services and in determining the sources of supply (including, in some cases, customers) for these services. Some services, such as system control and black start, will likely remain FERC regulated and continue to be priced under traditional cost-of-service rules.

System operators will play central roles in determining the amounts (and in some cases the locations) of ancillary services because (1) many of these services are needed for system reliability, (2) the system operator is often the only entity with sufficient information to decide what is needed where and when, and (3) it is often cheaper to provide a service for the aggregate of loads than for each load individually (i.e., economies of scale exist for services such as regulation).

During a multiyear transition period, as provision of some ancillary services shifts from an embedded-cost, one-size-fits-all paradigm to a market-based approach, some problems may arise. These problems will occur as certain suppliers, transmission owners, power marketers, and customers identify ways to "game" today's system in which prices are based on average costs and average loads rather than on specific cost-allocation factors.

As Stalon (1997) noted, developing reasonable governance structures for this new industry that appropriately balance commerce and reliability will be difficult. Thus, the establishment of reliability standards, such as those associated with minimum operating-reserve requirements, can no longer rely so heavily on rules of thumb and the judgment of system operators. In the future, with more and more participants in bulk-power markets, many of whom will be motivated primarily by profits, pressure will grow to provide a clear technical basis for these requirements and to quantify the tradeoffs between increased reliability and increased bulk-power trading. Developing such technical bases requires much more published data and analysis and may require national (rather than regional) reliability definitions and standards.

Bulk-power market participants might ask how much money could be saved by extending the maximum allowable recovery time from 10 to 15 minutes (which, presumably, would reduce the amounts of operating reserves required and their minimum ramp rates). How much more vulnerable would systems be to a second outage that might occur before the system was restored to balance? By how much would reliability be improved if the recovery time were shortened from 10 to 8 minutes?

Voltage control raises even more problems than do operating reserves because provision of voltage-control services (injection and absorption of reactive power) is much more location specific and can be provided by a variety of generation and transmission equipment. Perhaps because of these complexities, little effort has been made so far to find commercially efficient

ways of providing the resources needed for voltage control. FERC elected to separate generation- and transmission-based voltage control, establishing the former as an ancillary service while leaving the latter embedded in the basic transmission service. FERC also recognized that reactive power cannot be transported as far as real power, making it unclear if there will be a sufficient number of generators available to respond to locational control requirements to allow competitive markets to supply reactive power.

FERC made a reasonable start, but much more needs to be done. Market participants and their regulators need to agree on the appropriate mix of engineering standards (the approach chosen by the California ISO) and market mechanisms to acquire both the capability to produce and absorb VARs and the real-time delivery of this service. In particular, system operators may need to pay generators for any opportunity costs associated with a loss of real-power output to increase reactive-power output.

Ensuring that sufficient reactive resources are available may be increasingly difficult because of the deintegration of the electricity industry. Traditional vertically integrated utilities contained, within the same entity, generator reactive resources, transmission reactive resources, and the control center that determined what resources were needed when. In the future, these resources and functions may be placed within three different entities. In addition, these entities will have different, perhaps conflicting, goals. In particular, the owners of generating resources will be driven, in competitive generation markets, to maximize the earnings from their resources. They will not be willing to sacrifice revenues from the sale of real power to produce reactive power unless appropriately compensated. Similarly, transmission owners will want to be sure that any costs they incur to expand the reactive capabilities of their system (e.g., additional capacitors) will be reflected fully in the transmission rates they are allowed to charge.

State regulators can play important roles in helping bulk-power planning, operations, and markets to mature during the next several years. PUCs can work with their jurisdictional utilities in assessing alternative engineering and economic approaches to balancing reliability with commerce. They can advocate preferred solutions before FERC. And they can help educate state legislators on these difficult tradeoffs.

In some states, the PUCs may be able to exert a direct influence on the outcomes of these deliberations. Such states include those (1) with sufficient electricity production and/or consumption to affect regional markets (e.g., California), (2) in which history has focused electricity activities (e.g., New York with its one-state power pool and Texas, with its own interconnection and regional reliability council and independence from FERC), (3) in which tradition allowed state regulators to play important roles in bulk-power issues (e.g., New England with its New England Power Pool and the Pacific Northwest), or (4) with unique geography (e.g., Florida, which, although part of the Eastern Interconnection, is almost an electrical island). Much as the legislature and PUC did in California, these specially situated states can directly shape the structure and operations of their bulk-power markets.

Although FERC has primary jurisdiction over bulk-power issues, it has a clear policy of granting deference to states. In its recent order approving an ISO and power exchange for the PJM Interconnection, FERC (1997c) noted that its approval of the PJM Reliability Assurance Agreement is based in part on "the general preference of the state commissions within the PJM region that the traditional reliability aspects of the pool continue"

ACKNOWLEDGMENTS

We thank Kenneth Rose, the National Regulatory Research Institute project manager, for his continuing guidance and support throughout this project. We thank Carolyn Berry, Stephen Coe, Fred Heizer, Michel Hiser, Dan Johnson, Majid Khan, Steven Keller, James Loock, Kenneth Rose, Susan Stratton, and Jess Totten for their very helpful reviews of a draft of this report. We thank Fred O'Hara for editing the report.

REFERENCES

Atlantic City Electric et al. 1997, *Filing of the PJM Supporting Companies*, Docket No. EC97-38-000, before the Federal Energy Regulatory Commission, Norristown, PA, June 2.

T. Bilke 1997, personal communication, Wisconsin Electric Power, Milwaukee, WI, December.

M. Curley 1997, personal communication, North American Electric Reliability Council, Princeton, NJ, September.

D. Curtice 1997, *Cost of Providing Ancillary Services from Power Plants, Volume 1: A Primer*, EPRI TR-107270-V1, Electric Power Research Institute, Palo Alto, CA, March.

N. H. Dandachi, M. J. Rawlins, O. Alsac, M. Prais, and B. Stott 1996, "OPF for Reactive Pricing Studies on the NGC System," *IEEE Transactions on Power Systems* 11(1), 226-232, February.

East Central Area Reliability Coordination Agreement 1996, *ECAR Document No. 4, Criteria and Methods for the Uniform Rating of Generation Equipment*, Canton, OH, August.

East Central Area Reliability Coordination Agreement 1997, *ECAR/MAAC/SERC Tri-Regional Assessment Reliability Impact of the Delayed Completion of the Wyoming-Cloverdale 765-kV Line*, draft, Canton, OH, January.

Electric Reliability Council of Texas 1987, *Working Papers for the Group I Studies of the 1987 Spinning Reserve Task Force*, Austin, TX, July.

Electric Reliability Council of Texas 1997, *ERCOT Operating Guides*, Austin, TX, July.

A. A. El-Keib and X. Ma 1997, "Calculating Short-Run Marginal Costs of Active and Reactive Power Production," *IEEE Transactions on Power Systems* 12 (2), 559-565, May.

Florida Reliability Coordinating Council 1996, *FRCC Planning Principles and Guides, Principles and Guides for Planning Reliable Bulk Electric Systems*, Tampa, FL, September.

E. Hirst and B. Kirby 1996a, *Electric-Power Ancillary Services*, ORNL/CON-426, Oak Ridge National Laboratory, Oak Ridge, TN, February.

E. Hirst and B. Kirby 1996b, *Ancillary-Service Details: Regulation, Load Following, and Generator Response*, ORNL/CON-433, Oak Ridge National Laboratory, Oak Ridge, TN, September.

E. Hirst and B. Kirby 1997a, *Ancillary-Service Details: Dynamic Scheduling*, ORNL/CON-438, Oak Ridge National Laboratory, Oak Ridge, TN, January.

E. Hirst and B. Kirby 1997b, *Creating Competitive Markets for Ancillary Services*, ORNL/CON-4486, Oak Ridge National Laboratory, Oak Ridge, TN, October.

E. Hirst and B. Kirby 1997c, *Ancillary-Service Details: Operating Reserves*, ORNL/CON-452, Oak Ridge National Laboratory, Oak Ridge, TN, November.

ICF Resources 1997, *Independent Transmission System Operators and their Role in Maintaining Reliability in a Restructured Electric Power Industry*, Fairfax, VA, September.

Interconnected Operations Services Working Group 1997, *Defining Interconnected Operations Services Under Open Access*, EPRI TR-108097, Electric Power Research Institute, Palo Alto, CA, May.

J. P. Kessinger 1997, "Flexible AC Transmission Systems," presented to DOE Task Force on Electric-System Reliability, Plymouth, MA, September 24.

D. Kincheloe and R. E. Burns 1997, "The Continuing and Evolving State Commission Role in Assuring Electric Reliability," *NRRI Quarterly Bulletin* 18(2), 179-182, Summer.

B. Kirby and E. Hirst 1996, *Ancillary-Service Costs for 12 U.S. Electric Utilities*, ORNL/CON-427, Oak Ridge National Laboratory, Oak Ridge, TN, March.

B. Kirby and E. Hirst 1997, *Ancillary-Service Details: Voltage Control*, ORNL/CON-453, Oak Ridge National Laboratory, Oak Ridge, TN, December.

P. Nedwick; A. F. Mistr, Jr.; and E. B. Croasdale 1995, "Reactive Management: A Key to Survival in the 1990s," *IEEE Transactions on Power Systems* 10(2), 1036-1043, May.

New England Power Pool 1996, *New England Power Pool (NEPOOL) Restructured Arrangements (including Open Access Transmission Tariff)*, submitted to the Federal Energy Regulatory Commission, Holyoke, MA, December 31.

New York Power Pool 1997, *Comprehensive Proposal to Restructure the New York Wholesale Electric Market*, submitted to the Federal Energy Regulatory Commission, Albany, NY, January 31.

North American Electric Reliability Council 1995, *NERC Operating Manual*, Princeton, NJ.

North American Electric Reliability Council 1996a, *Policy 2, Transmission Operations & Voltage and Reactive Control*, Princeton, NJ, October.

North American Electric Reliability Council 1996b, *Performance Standard Training Document*, Princeton, NJ, November 1.

Northeast Power Coordinating Council 1992, *Guidelines for Inter-AREA Voltage Control*, NPCC Operating Procedure Coordinating Committee and NPCC System Design Coordinating Committee, New York, NY, December.

Northeast Power Coordinating Council 1996, *Memorandum of Agreement*, New York, NY, November.

Pacific Gas and Electric, San Diego Gas & Electric, and Southern California Edison 1997, *The Phase II Filing of the California Independent System Operator Corporation and The Phase II Filing of the California Power Exchange Corporation*, submitted to the Federal Energy Regulatory Commission, Docket Nos. EC96-19-001 and ER96-1663-001, San Francisco, San Diego, and Rosemead, CA, March 31.

A. Panvini and T. J. Yohn 1995, "Field Assessment of Generator Reactive Capability," *IEEE Transactions on Power Systems* 10(1), 288-296, February.

PJM Interconnection 1997, *Dispatching Operations Manual M-12*, Revision 01, Norristown, PA, July.

Southeastern Electric Reliability Council 1995, *Principles and Guides for Reliability in System Planning*, Birmingham, AL, April.

Southwest Power Pool 1997, *Southwest Power Pool Independent System Operator Proposal*, Little Rock, AR, March.

C. G. Stalon 1997, "Electric Industry Governance: Reconciling Competitive Power Markets and the Physics of Complex Transmission Interconnections," *Resource and Energy Economics* 19(1-2), 47-83, March.

U.S. Federal Energy Regulatory Commission 1996, *Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, Final Rule*, Docket Nos. RM95-8-000 and RM94-7-001, Order No. 888, Washington, DC, April 24.

U.S. Federal Energy Regulatory Commission 1997a, *Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities*, Order on Rehearing, Docket Nos. RM95-8-001 and RM94-7-002, Order No. 888-A, Washington, DC, March 4.

U.S. Federal Energy Regulatory Commission 1997b, *Opinion and Order Affirming in Part and Reversing in Part Initial Decision*, Opinion No. 416, Docket Nos. ER91-150-000 et al., Washington, DC, September 16.

U.S. Federal Energy Regulatory Commission 1997c, *Pennsylvania-New Jersey-Maryland Interconnection: Order Conditionally Accepting Open Access Transmission Tariff and Power Pool Agreements, Conditionally Authorizing Establishment of an Independent System Operator and Disposition of Control over Jurisdictional Facilities, and Denying Rehearings*, Docket Nos. OA97-261-001 et al., Washington, DC, November 25.

S. Virmani, E. Lo, and D. McNair 1992, *Impacts of Governor Response Changes on the Security of North American Interconnections*, TR-101080, Electric Power Research Institute, Palo Alto, CA, October.

Western Systems Coordinating Council 1997, *WSCC Operating Reserve White Paper*, Salt Lake City, UT, June 4.

D:\Wpd\ANC\c454.wpd