EDISON ELECTRIC INSTITUTE

POWER MARKET AUCTION DESIGN

Rules and Lessons in Market-Based Control for the New Electricity Industry

PREPARED BY

Mathew J. Morey
Envision Consulting
Alexandria, VA

PREPARED FOR

EDISON ELECTRIC INSTITUTE

SEPTEMBER 2001

Edison Electric Institute (EEI) is the association of U.S. shareholder-owned electric companies, international affiliates and industry associates worldwide. Our U.S. members serve nearly 95 percent of all customers served by the shareholder-owned segment of the industry. They generate almost 70 percent of all the electricity generated by electric companies in the country and service about 70 percent of all ultimate customers in the nation.

Organized in 1933 and incorporated in 1970, EEI works closely with its members, representing their interests and advocating equitable policies in legislative and regulatory arenas. In its leadership role, the Institute provides authoritative analysis and critical industry data to its members, Congress, government agencies, the financial community and other influential audiences. EEI provides forums for member company representatives to discuss issues and strategies to advance the industry and to ensure a competitive position in a changing marketplace.

EEI's mission is to ensure members' success in a new competitive environment by:

- Advocating Public Policy
- Expanding Market Opportunities
- Providing Strategic Business Information

For more information on EEI programs and activities, products and services, or membership, visit our Web site at www.eei.org.

To place a publications order, call **1-800-EEI-5453**.

EDISON ELECTRIC
INSTITUTE

Power Market Auction Design: Rules and Lessons In Market-based Control for the New Electricity Industry

Mathew J. Morey

Envision Consulting
Alexandria, VA 22314-4813

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

August 2001

Published by:
Edison Electric Institute
701 Pennsylvania Avenue, NW
Washington, D.C. 20004-2696
Phone: 202-508-5000
Web site: www.eei.org

CONTENTS

List of Acronyms

AC	Alternating current
ACE	Area control error
AGC	Automatic generation controls
CalPX	California Power Exchange, Inc.
CFD	Contract for differences
CP	Clearing price
Dec	Decremental adjustment bid
ERCOT	Electric Reliability Council of Texas
FCC	U.S. Federal Communications Commission
FERC	U.S. Federal Energy Regulatory Commission
FTR	Financial transmission right
Inc	Incremental adjustment bid
ISO	Independent System Operator
ISO-NE	Independent System Operator of New England, Inc.
LMP	Locational marginal price
LOLP	Loss of Load Probability
LSE	Load-serving entity
MCP	Market clearing price
MSS	Multi-settlement System
NEPOOL	New England Power Pool
NERC	North American Electric Reliability Council
NYISO	New York Independent System Operator
NYSEG	New York State Electric and Gas Company
PAB	Pay-as-bid
PJM ISO	Pennsylvania-New Jersey-Maryland Interconnection, L.L.C.
PPA	Power Purchase Arrangement
QSE	Qualified Scheduling Entities
RMR	Reliability Must Run
RTO	Regional Transmission Organization
SO	System Operator
SMP	System Marginal Price
VOLL	Value of Lost Load

Summary

This paper takes a look at energy auction design options and problems and the experiences of regional transmission system operators with energy auction designs. The increasing reliance on competitive wholesale and retail electricity markets has led to widespread use of auctions to price and allocate energy and energy-related services. The primary goal in designing an energy auction should be the lowest-cost dispatch of generation that balances supply and demand and minimizes transmission congestion and system contingency management costs. This goal should be achieved through an open, transparent competitive process. A transparent auction yields a fair process in which the market provides much of the regulatory oversight.

Auctions are organized markets where goods are awarded to bidders based on rules that determine who wins and the price the winner pays. Auctions can be used to sell products and property rights (*e.g.,* wine, art work, or the right to drill for oil in the Gulf of Mexico or a right to use the transmission system when it becomes congested) or to award contracts to potential suppliers (*e.g.,* for road construction projects or, in the case of electricity, contracts to supply energy to the distribution utility purchasing on behalf of residential consumers).

From the perspective of regulators and antitrust agencies dealing with competition policy, the most important issues in auction design for energy markets may well be the traditional concerns of preventing collusive, predatory, and entry-deterring behavior. The regulatory objective should be to promote designs that suppress gaming or render it ineffective in favor of greater efficiency. The principle is to treat the market design as establishing a *mode of competition* among market participants. The trick is to select the mode of competition that most effectively realizes the potential gains from trade.

One caveat is in order: good auction design cannot offset structural imperfections that reduce competitiveness and contribute to market power problems. Short-run market efficiency may be reduced by physical or technical limits on the sellers' or buyers' ability to respond to changes in the short-run market price.

The Goal of Auction Design: Economic Efficiency

The most important reason for caring about auction design is that, without careful consideration for designs whose rules are complete, consistent, and without loopholes, energy markets will be less efficient and the objectives of the auction and the best intentions of regulators will not be fully realized. A poor auction design (one with rules that are incomplete, inconsistent, or full of loopholes) may result in prices of energy, ancillary, and transmission services that do not reflect sellers' marginal costs or buyers' marginal willingness to pay, something prices would normally represent if they were set by a competitive (efficient) market.

The efficiency of the auction is in large part determined by its rules. Auction rules (also known as activity rules or activity protocols) determine how the auction proceeds: who may participate, whether bids must exceed a particular threshold to be viable, whether changes to bids will be allowed, how large a bid increment must be, when a round of bidding stops and the next begins, and so forth.

Effective auction designs encourage short-term and long-term efficiency by suppressing gaming[1] or rendering it ineffective and establishing incentives for behavior by market participants' that are consistent with their objectives. For example, rules that make bids binding after a particular time in the auction, that assess penalties for scheduling changes or limit changes to particular periods, accompanied by strong monitoring and strict compliance enforcement, encourage market participants to truthfully represent their market positions, *i.e.*, reveal their true opportunity costs in their bids and generation commitment schedules.[2] Auction rules also increase market efficiency by enabling power markets to be contested by lowering entry and transaction costs or reducing uncertainty about the probability of winning, especially for suppliers with smaller asset bases and more limited financial resources.

Energy Market Fundamentals

Trades between buyers and sellers can be arranged in two basic ways. Buyers and sellers can pair up—one buyer with one seller—and reach agreement on the terms of exchange (otherwise known as a "bilateral" trade), or sellers can sell their product to an intermediary (a centralized or mediated market, for example) who, in turn, sells it to final consumers. Bilateral and mediated markets come in several varieties, bilateral markets being generally less organized, but there is some overlap between the types.

The most salient distinction of bilateral energy markets is the continual process of trading, with prices unique to each transaction. In contrast, mediated energy markets (power pools and power exchanges) are typified by a uniform price that all buyers pay and all sellers receive, with an auction run at regular times to set the market price in advance of physical delivery.

The long-term energy market, where long-term refers to future periods of a week or longer, up to many years, naturally functions as a combination of both mediated and bilateral markets. Buyers and sellers can search each other out and agree in private on the delivery of specified quantities of power at some specific times in the future through a customized contract called a *forward*. Or they can turn to a central exchange, called a *futures* market, in which standardized (regulated) contracts are traded.

1 Gaming is defined as a market participant's strategic behavior designed to profit from exploitation of recognized gaps or deficiencies in market structure and market rules, *e.g.*, bidding strategy in a simultaneous, multi-round auction for energy and capacity that might take advantage of information revealed in early rounds about the market positions of other suppliers.

2 Opportunity cost is an economist's phrase for the value of the most desirable alternative given up when choosing an option or use of a resource. For example, choosing to buy a car may mean that no art work can be purchased for a year or more.

The shorter-term energy market, where shorter-term refers to a future of a day or less, down to hours or minutes before specific generators are dispatched, becomes less amenable to combinations of market types, and more amenable to mediated market forms such as exchanges and pools run as auctions. As the time to actual delivery draws near, the opportunity cost to search for an appropriate bilateral trade increases and factors affecting consumption and production become more certain, so buyers and sellers tend to find mediated markets more economical for trading, although bilateral trades are not ruled out entirely.

Exchanges and pools (*i.e.*, auctions) are naturally efficient trading forms when the product traded is integral to the functions and services regularly performed by the transmission system operator (SO) and quantities of various products likely are not well known until about a day before physical trading takes place. The SO operates as an intermediary to purchase and provide ancillary services for transmission users who do not buy them directly. The SO typically relies on an auction to efficiently obtain the needed quantities in a shorter-term market, such as a day-ahead market.

Price setting in competitive energy markets adheres to economic principles that dictate outcomes in other competitive commodity or product markets: all winning buyers' valuations of the commodity are equal to or higher than the price and all winning sellers' costs of production are less than or equal to the price. Energy auctions (pools and exchanges) can be designed to set a single uniform market-clearing price or to set multiple prices for a commodity, but the principles still apply. The details of these mechanisms are discussed in Section 4. In a competitive auction or bilateral market, the price that satisfies the aggregate demand of winning buyers and covers the aggregate costs of the winning sellers is called the market-clearing price, in the jargon of the electricity industry. This price clears the market in that the aggregate quantity demanded (the sum of all individual buyers' quantities demanded) and aggregate quantity supplied (the sum of all the individual sellers' quantities supplied) are equal, a condition referred to by economists as market equilibrium.[3] If the market price were set lower, there would be more quantity demanded than would be supplied, and if it were set higher, there would be less quantity demanded than supplied. In other words, the market would not be in equilibrium, that is, it would not clear.

3 Economists often express the equality of aggregate supply and demand in terms of marginal costs and values. Thus, the market-clearing price equates the marginal cost of producing the last unit of the commodity demanded and the marginal value of the last unit of the quantity supplied.

Auction Basics

Auction designers have four basic types of auction formats from which to choose that have been widely analyzed and applied:

- the ascending-bid auction (also called the open, oral, or English auction),
- the descending-bid auction (used in the sale of flowers in the Netherlands and so also called the Dutch auction by economists),
- the first-price sealed-bid auction, and
- the second-price sealed-bid auction (also called the Vickrey auction by economists).

All four basic auction forms can lead to a *uniform* winning price when a single unique object or good is auctioned. However, energy auctions involve the sale of multiple indistinguishable units of an object, for example, shares of load measured in MWh, for a given hour of the day, or shares of generation capacity measured in MW, to supply backup or replacement reserves in case of system contingencies for satisfying system reliability standards. In these cases, an alternative to a uniform price that does not discriminate among winning bidders, would be to discriminate by running a first-price sealed bid auction in which each winning bidder pays her bid, or is paid her offer price as in the case of an energy procurement auction run by a system operator. Further discussion of the merits of uniform price vs. pay-as-bid price auctions is taken up in Section 7.

Economists have shown theoretically that the outcomes of any of the four basic auction forms will be equivalent under particular assumptions.[4] This theoretical result, called the revenue equivalence theorem, makes two extremely important statements about energy or ancillary service auctions. First, it says that the total cost to society of procuring energy or ancillary services through any of these auction forms will be the same regardless of form. Second, and what may be most important for policy makers, it says that the costs are independent of the form because buyers' and sellers' bidding strategies will be adjusted as the auction rules change.

Well-designed auctions follow a set of fairly simple, sensible principles. The goals of the auction must be clearly established, the objects to be auctioned well defined, and the rules of the auction written so that serious players are encouraged to participate and their behavior is consistent with a competitive market. Activity rules help control bidding by dictating the details of the bidding in each round of an auction, such as the form of bid submissions, the timing of bids, the size of bid increments in a multi-round auction, or the manner in which a bidder can change a bid or exit the auction.

4 *Myerson* and *Riley* and Samuelson.

Markets for Energy, Capacity, and Reserves

The key to efficient wholesale power markets is the design of the auction mechanisms for facilitating trades or allocating rights to resources (capacity or energy) and transmission rights. All working models for creating new competitive electricity markets include a single network operator responsible for controlling the physical operation of a control area, coordinating generator schedules, balancing loads and supply resources in real time, acquiring ancillary support services required to maintain reliability and coordinating with neighboring control areas. However, there remains considerable disagreement regarding the scope of the system operator's involvement in the energy and ancillary markets. The vast majority of the real-world applications of the hybrid pool/bilateral market model have the system operator administering day-ahead forward and near-real-time auctions for energy or capacity or both on a daily basis. In most cases, bilateral trades are permitted which can also constitute a longer-term forward market, and in California until 2001, there was a separate exchange that administered day-ahead and longer-term forward contract markets.

In sum, the case for a poolco or centralized auction design is strongest when there is vigorous competition and good optimization. The case for decentralization, *i.e.*, a bilateral forward market, is strongest when tight coordination in forward markets is less important than good scheduling decisions by each participant, provided, of course, that a system operator manages the transmission system in real time and conducts a liquid real-time market. Whether the deficiencies of centralized designs used in practice are more important than incompleteness of those decentralized markets feasible in practice depends on the situation (and changes as technology changes) and depends crucially on their comparative advantages in controlling abuses of market power and stimulating competition and entry.[5]

Constraints on Energy Auctions

Energy auctions are frequently conflicted because they typically assume inter-temporal independence in production costs when setting up and running day-ahead hourly auctions. However, constraints on generators pose serious challenges to auction designers seeking to mitigate the opportunities these constraints create for suppliers to profit from strategic bidding. Forward market auction designs must account for how generator bids and information on unit constraints will be used by the system operator in determining unit commitments and the least-cost dispatch for each period of the following day. The complexity of the unit commitment problem and the imperfections in computing software may enable generators to manipulate their bids to ensure particular units are included in the supply stack, perhaps at prices above competitive market-clearing levels. To make the task even more complicated, often revisions to bids or unit schedules are permitted after all clearing prices are revealed, creating incentives for generators to adjust their schedules—to change the units that will be operated in real time to serve a particular market so as to take

---

5 For further discussion of these issues, *see Wilson* 1999.

advantage of periods where clearing prices are high and supplies are inadequate to meet demand.

Complete designs for energy and some ancillary service markets would accommodate the demand as well as supply in a double-sided auction. Much debate has ensued about how best to incorporate the demand side into wholesale power markets. What can be achieved will be determined by various factors that limit voluntary participation by end-use customers or load-serving entities (LSEs) through offers to reduce load. Part of the challenge in integrating load bidding in the wholesale markets stems from the great diversity of end use technologies, usage patterns, and the value of end use appliances and technologies to consumers. The diversity would appear to be significantly greater than what is seen on the supply side. A major constraint at the moment is technical: for load to participate directly in forward and real-time markets it must be metered appropriately and in some cases, directly controllable (*i.e.*, dispatchable) by the system operator. According to *Hirst*, "dynamic-pricing and load-reduction programs require interval meters, communication systems to move data and instructions between the customer and its LSE, and perhaps automatic-control systems that respond to time-varying prices."

Settlement Systems

Settlement systems, the rules concerning the price(s) that will be paid to suppliers or paid by buyers, pose another challenge for auction designers. Designers can choose between single- or multi-settlement systems. The single-settlement system may appear simpler than a multi-settlement system because it involves a single set of hourly prices and is closer to the way tight power pools operated before wholesale restructuring. However, this simplicity is deceptive. The difficulty with the single *ex post* settlement is that much is riding on the *ex post* prices, since all earlier commitments and transactions are settled at the prices established in real time. After the day-ahead schedule is formed, bidders have an incentive to make adjustments to influence the real-time price in a favorable direction. Since the real-time price is used for all trades, the incentive for manipulation may be large.[6] For instance, day-ahead transactions, including bilateral transactions, may account for 95 percent of trades, but these are settled at prices that reflect heavily the 5 percent traded in the real-time market. Bidders can take advantage of short-term inelasticities in the supply and demand schedules to reap greater profits. Knowing how to do this is complex, and can be exploited best by large bidders with sufficient scale (*i.e.*, larger, more diverse generator portfolios) to make the

6 Contracts for differences (CfDs, bilateral hedging contracts) may make the real-time price irrelevant for a significant portion of transactions. However, other bilateral transactions may be tied to the real-time price. Also the spot price may indirectly influence the terms negotiated in bilateral contracts. The seller bids supplies into the energy market and the buyer purchases from the energy market. The seller receives the spot price; the buyer pays the spot price. However, the CfD has a strike price known only to the contract participants. If the strike price is above the spot price for the hour, the buyer pays the seller the difference; if the strike price is below the spot price, the seller pays the buyer the difference. The net payment is the contract strike price.

efforts worthwhile. The added complexity and risk tends to discourage entry and participation by small bidders whose net revenue might be whipsawed by price volatility in the real-time market. This gaming can be mitigated by financial penalties for failures to perform as scheduled, but this raises the question: what are the appropriate penalty values?

A multi-settlement system mitigates gaming on two fronts. First, the day-ahead bids are binding financial commitments. The bids and resulting schedules are credible precisely because they are financially binding. Second, bidders are unable to alter the day-ahead prices. Day-ahead prices remain fixed for all transactions scheduled in the day-ahead market. Deviations from the day-ahead schedule affect the real-time price, but the real-time price is used only to price these deviations. Hence, in a multi-settlement system, the incentive to manipulate the real-time price is not magnified as it is in a single-settlement system. Penalties for non-performance are not needed in a multi-settlement system, since deviations from the schedule are priced correctly. If a generator fails to deliver as scheduled, it is liable for the quantity it was supposed to deliver priced at the real-time price.

A debate has arisen among economists and market participants over the benefits of conducting a uniform price (non-discriminatory) versus a second-price or pay-as-bid (PAB) (discriminatory) auction.[7] The arguments can be generalized into the distinctions made between non-discriminatory and discriminatory auction designs. Auction theory has identified two effects on suppliers' bidding behavior that make it difficult to know in advance which design is best. The presence of the winner's curse argues for a uniform price auction while the influence of inframarginal capacity on market-clearing prices argues for a PAB auction. In either case, the strategies of suppliers will depend, in part, on how many bidders participate in the auction, in other words, will depend on the degree of competition in the market, and how much information each of the suppliers has relative to competitors. Whether the winner's curse or inframarginal capacity has the greater effect on the level of market-clearing prices is probably a function of specific attributes of the energy market. It is by no means clear which effect will be stronger in energy auctions. Perhaps only time will reveal the answer.

Auction Designs in US Regional Power Markets

Power market auction designs have been developed to fit unique regional circumstances. Circumstances that shaped the design of the existing regional power markets have included the vestiges of vertical integration (*e.g.*, native load priority for transmission capacity), close regulatory oversight of retail market prices that insulated customers from wholesale price volatility, continued obligations for distribution utilities as providers of last resort and standard offer service, vested interests in revenue neutrality (*i.e.*, revenue preservation), and concerns about cost shifting and generator market power, especially in regions of the country subject to a transition to competition at the retail level. These issues

7 *See* Section 7.4 for further details.

helped shape the auction designs for particular product markets such as real-time balancing, forward energy, or ancillary services.

Designs also vary because the process that produces them is highly politicized. The process attempts to balance the disparate interests of various groups of market participants. At the wholesale level, this is partly a consequence of the insistence in Order No. 2000 on RTO governance structures that are independent of market participants. However, market participants still have a voice in market design by their participation on stakeholder committees that offer advice to the governing boards of the regional system operators.[8] This form of governance structure, generally sanctioned by FERC, complicates the decision making for design because the outcome often reflects a compromise that makes stakeholders happy but sacrifices design features that would make the auction or the market more efficient.

The ISOs' real-time market auction designs are similar: bid and offer schedules are obtained in a day-ahead, double-sided auction used for scheduling and determining unit commitments for the next day. For balancing, bids and offers are submitted as incremental and decremental adjustments to scheduled quantities, with conditions on how long such adjustments can be sustained before restoration to original levels. Settlements are based on a uniform market-clearing price computed from either real-time locational marginal prices (*e.g.*, as in PJM or New York and in New England by the end of 2001) or from zonal prices (*e.g.*, as in California).[9]

Settlements in each of these markets are based on some variation of a uniform real-time hourly clearing price, computed at a system bus or averaged over a zone. In New England, the hourly settlement price is calculated *ex post* as a weighted average of real-time marginal cost prices during each hour. In New York and PJM, settlements are based on local marginal prices at nodes (*i.e.*, busses) within the region calculated as frequently as every 5 minutes. Table 4 summarizes the main features of the real-time balancing markets run by the ISOs.[10]

---

8 As *FERC* states in Order No. 2000, ". . . even among the ISOs, there are different models of governance. As we noted in the NOPR, the dominant governance model (PJM, New England, New York and the Midwest) for ISOs is a two-tier form of governance. The top tier consists of a non-stakeholder board, while the lower tier consists of advisory committees of stakeholders that may recommend options to the non-stakeholder board. Generally, the top tier has the final decision making authority.

9 Zonal prices are typically computed as some weighted average of locational market-clearing (*i.e.*, marginal) prices.

10 Definitions of acronyms used in the Tables 4, 5 and 6 are: AS = ancillary services; avg = average; Cap = capacity; CP = clearing price; DA = day ahead; HA = hour ahead; imb = imbalances; inc = incremental; LMP = locational marginal price; MCP = market-clearing price; mkt = market; MSS = multi-settlement system; OMO = out of merit order; op. res. = operating reserves; QSE = qualified supplier of electricity; RT = real time; sched. Coor. = scheduling coordinators; TMSR = ten-minute spinning reserve; TMNSR = ten-minute non-spinning reserve; "TMR" = thirty minute reserve;

Basic features of the day-ahead energy markets in the regional systems are summarized in Table 5. The day-ahead energy markets in New York and California have experienced similar problems of under-scheduling by buyers in an attempt to influence prices in the day-ahead market and obtain lower prices in the real-time market. In New York, the incentive to under-schedule in the day-ahead market was a function of a rule that socialized a portion of the ISO's costs of meeting imbalances in the uplift charges allocated to all users. This rule has been changed to shift under-scheduling cost responsibilities to the buyers.

The original designs of the ancillary services procurement auctions administered by each of the regional transmission operators display considerable diversity, although there has been a convergence to a common model over the past two years. Each of the FERC-approved ISOs provides the ancillary services required under Order No. 888. All transmission customers are required to purchase energy balancing service from the ISO. Ancillary services such as spinning reserves, non-spinning reserves, and replacement reserves can be self-provided and self-scheduled, obtained through bilateral contracts or obtained from the ISO. However, the ISOs are required to be providers of last resort for these services, and they generally provide them to a large percentage of their regional market's transmission customers.

Most of the ISOs' original ancillary service auction designs experienced similar problems for a variety of reasons. For example, price reversals and increasing out-of-merit-order reserve service costs were experienced in all ISO markets and often wound up in the ISO's uplift or operating costs allocated to all transmission system customers. The ability of suppliers to obtain higher prices for lower quality ancillary services was a combination of at least three things: activity rules, transmission congestion limiting supply response to higher prices, and limited or non-existent short-run demand response. In addition, some of the problems stemmed from the software design or from unrestricted sequential auctions, as in California. Consequently, the designs have undergone revisions over the past two years and have been converging to the rational (or smart) buyer model described in Section 4.2.4.

In each of the three ISOs in the Eastern Interconnection, an installed capacity market has been established that pays generators for reserving capacity to meet reliability requirements on the system. The LSEs are required to have or contract with generators for a prescribed level of reserve capacity above their peak load within a certain time frame. Formal or informal secondary capacity markets that allow trading of capacity obligations among the LSEs have accompanied reserve capacity obligations.

The reserve requirements and capacity markets provide generators with an opportunity to collect marginal revenue for their unutilized reserve generation capacity. Capacity payments also provide incentives for the building of reserves beyond those sufficient to meet short-term ancillary services needs.

The need for a capacity market and reservation payments stems largely from the fact that demand in all of these markets is highly price inelastic in the short term and does not participate directly in the short-term wholesale markets. One problem with capacity payments is that they are often set equal to the value of peaking technology capacity costs, thus generators are fully compensated even if they sit idle. Consequently, such payments may induce inefficiently high investment in capacity.

Recommendations

If policy makers hold economic efficiency out as one of the objectives in restructuring the wholesale and retail electricity markets, then, to restate the obvious, market and auction designs matter. Designs, good or bad, will create modes of competition among buyers and sellers that will yield more or less efficient prices, resource uses and allocations for future use. The challenge for policy makers is to provide the framework that will enable experts in design to create auctions that can induce behavior consistent with that efficiency objective.

Federal and state policy on market and auction design should be based on a technical/economic framework, built upon the work of leading market designers and the experience with designs to date. Given that the rest of the country has yet to develop full-blown regional wholesale power and transmission markets under Order 2000, and all of these markets ultimately will have to be seamlessly integrated, the time is ripe for such a document to provide much needed guidance to nascent regional transmission organizations.

The policy framework should encourage incentive compatible behavior through designs that:

- Avoid socialization of generation costs in uplift charges that encourages gaming by both sellers and buyers and masks the true costs of generation at the margin in real time, sending inefficient price signals to both sides of the market.
- Closely synchronize the activity and bidding rules for short-term (*i.e.*, day-ahead) energy and ancillary service markets when they are mediated by separate agencies, such as by an independent exchange for energy and by the system operator for ancillary services.
- Examine market conditions to determine whether proposed PAB designs would be superior to uniform price designs: to see whether the winners' curse or inframarginal capacity is likely to have greater effect.
- Introduce demand-side bidding in day-ahead energy and ancillary service auctions to discipline supply and to lead to lower reliance on or elimination of capacity markets, paying close attention to align bidding rules with the factors that motivate buyers.
- Expand on the use of multi-settlement systems and eliminate reliance on single-settlement designs to help eliminate gaming.
- Expand on the use of multi-part bid designs to better reflect opportunity costs in supplier and buyer offers and bids.
- Restrict price bids in sequential auction formats to be consistent with the rational or smart buyer model, to prevent price reversals in ancillary service markets, or

better yet, rely on simultaneous auction formats for energy and ancillary services. and

- Keep in mind the "revenue equivalence theorem" that implies buyers and sellers will adjust their bid and offer strategies to fit the auction design so that the total cost of the auction will be equivalent regardless of the design choice.

The most important issues in auction design for energy markets may well be the prevention of collusive, predatory, and entry-deterring behavior. The regulatory objective should be to put in place policies that promote designs that suppress gaming or render it ineffective in favor of greater efficiency. Identifying the best locus of incentives and competitive forces becomes the central design problem and the greatest challenge for designers. The complexity of the network and its operations, the interdependency of generation and transmission products and services, and the vast number of design options ensures that market and auction designs will be less than perfect the first time. Thus, design will be an iterative process and designers may need to build in flexibility for performance review and design changes based on actual experience.

Part I ✥ Introduction

This paper takes a look at energy auction design options and problems and the experiences of regional transmission system operators with various energy auction designs. The increasing reliance on competitive wholesale and retail electricity markets has led to widespread use of auctions as a means to price and allocate energy and energy-based ancillary services. The vast majority of electricity is exchanged through the use of private bilateral contracts, as high as 80 to 90 percent in some regional markets. However, a small, but nonetheless significant, portion is exchanged through energy auctions, especially for critical reliability and system balancing functions. This means that the auction design can have a significant effect on the efficiency of the entire market for energy and related services.

Widespread application of auctions to electricity markets is a relatively recent development, a consequence of the broader industry restructuring that typically separates generation from transmission and distribution. Therefore, auctions, in general, are not widely understood by market participants or regulators, and energy auctions even less so. Electricity physics and network operations make designing efficient electricity auctions more complicated than for other commodities, more vulnerable to gaming and manipulation, and, therefore, more challenging even for expert auction designers. However, a well-designed auction can play an important part in making competitive electricity markets efficient.

The primary goal in designing an energy auction should be the lowest-cost dispatch of generation that balances supply and demand (load) and minimizes transmission congestion and system contingency management costs. The auction should produce the lowest cost, reliable short-term and long-term sources of electricity available in the current market.

From a regulatory policy perspective, a second goal should be that an open and transparent competitive process be used to achieve the primary goal. A transparent auction yields a fair process in which the market provides much of the regulatory oversight. Fortunately, these goals are not in conflict. A well-designed auction encourages the open competition that is the clearest sign that the best possible terms have been found for energy and ancillary services.

However, the very best auction designs can still suffer from inefficiencies connected to the misalignment of rules and participants' private incentives. The misalignment can be the consequence of physical or technical design limitations arising from network operations, the physics of electricity flows or the structure of supply and demand. Furthermore, electricity market designers and policy makers have begun to realize an age-old law: behavior changes to accommodate changes in the rules. As *Chao* and *Wilson* have stated:

> Gaming strategies are inherent in any design that requires traders to manipulate their bids in order to take account of factors that the bid format does not allow them to express directly.

Auctions are organized markets where goods are awarded to bidders based on rules that determine who wins and the price the winning bidder pays. Auctions can be used to sell products (*e.g.,* wine, art work, or the right to drill for oil in the Gulf of Mexico or a right to use the transmission system when it becomes congested) or to award contracts to potential suppliers (*e.g.,* for road construction projects or, in the case of electricity, contracts to supply energy to the distribution utility purchasing on behalf of residential consumers).

Auctions are an age-old market form, dating back centuries, but have become extremely popular over the past decade as a consequence of advances in auction theory made in the field of operations research, applications of auctions in experimental testing of market designs, and the explosion of applications arising from growth in e-commerce that has encouraged businesses to organize auctions for many inputs and outputs whose trade was previously negotiated bilaterally. For example, auctions have been applied worldwide to allocate the radio spectrum. Auctions now are the principal means for organizing centralized electricity exchanges (called pools) and other related electricity product markets, such as ancillary services that support transmission system reliability.

Economists have been enormously successful at promoting auction theory, but the resulting practice has been less successful. The US wholesale power market auction experience provides lessons about the importance of design to the performance of a competitive electricity market. The experiences over the past two years in some regional electricity markets (*e.g.*, with price volatility, price spikes, and increasing costs to maintain system reliability) can be attributable, in part, to designs that are incompatible with the economic incentives of market participants. Today's regional wholesale power market auction designs are less efficient because the rules motivate participants to behave in ways inconsistent with the social objective of lowering overall costs of electricity generation and delivery. Corrections are being made to the designs that should improve performance of the energy and ancillary service markets.

From the perspective of regulators and antitrust agencies dealing with competition policy, the most important issues in auction design for energy markets may well be the traditional concerns of preventing collusive, predatory, and entry-deterring behavior. The regulatory objective should be to promote designs that suppress gaming or render it ineffective in favor of greater efficiency. The principle, however, is to treat the market design as establishing a *mode of competition* among the traders. The trick is to select the mode of competition that is most effective in realizing the potential gains from trade. Identifying the best locus of incentives and competitive forces becomes the central design problem, and the biggest challenge for designers who often may not have complete information about the factors that shape market participants' behavior. Consequently, market and auction design may be necessarily an iterative process and designers may need to build in flexibility for performance review and design changes, based on actual experience; one should not expect that the designs will be perfect the first time out of the box.

Recommendations about optimal auction design for power and transmission rights markets and the activity rules for them have been and will continue to be based on economists' applications of game theory to ensure the outcomes of the auctions are efficient allocations of scarce resources. Several parts of economic theory prove helpful in designing the auction rules and in thinking about how designs might be improved and adapted.

However, good auction design, for the most part, consists of good elementary economic theory coupled with a good dose of common sense about what motivates economic agents.

One caveat is in order: good auction design cannot offset structural imperfections that contribute to market power problems.[11] Short-run market efficiency may be reduced by physical or technical limits on the sellers' or buyers' ability to respond to changes in the short-run market price. For example, network congestion can reduce the number of suppliers able to reach a local market by some transmission path and give local suppliers an opportunity to raise the market prices without suffering competitive consequences. Another example is where large-volume industrial consumers do not see the real-time variation in the price of electricity, and hence the cost to them of consuming at particular times; they are not able to respond to higher prices by reducing consumption. Price responsive consumers can contribute to reducing price volatility and to imposing greater discipline on suppliers' bidding behavior. No amount of good auction design will correct these types of imperfections that may inhibit market efficiency and other steps may have to be taken in the short-run to sharpen performance.

This paper is organized as follows. Section 2 discusses how the objectives of auction design are important for achieving market efficiency. Section 3 explains the organization of markets and the relationships of market forms to the various energy and related products that will be traded through an interconnected network. Section 4 lays out the fundamentals of auctions and principles for auction design. Section 5 discusses the various product markets for energy, ancillary services and capacity, their interdependencies and the challenges they pose for designing effective auctions. Section 6 discusses factors on the supply and demand sides that constrain auction designs. Section 7 defines single and multi-settlement systems and their pros and cons. Section 8 summarizes the designs for balancing, energy, ancillary and capacity market auctions now operating in the regional markets administered by independent system operators. Section 9 offers conclusions and some recommendations for a new direction in policy on market design.

11 A distinction needs to be made between market design—having to do with the *institutional form(s) of the markets* and the rules that govern participant behavior within them—and the *structure of the markets*—having to do with the number buyers and sellers, the numbers and types of products exchanged, the ease of access to information and the transparency of the market price.

Part 2 ❖ The Objective of Auction Design: Economic Efficiency

The most important reason for caring about auction design is that, without careful consideration for designs whose rules are complete, consistent, and without loopholes, energy markets will be less efficient and the objectives of the auction and the best intentions of regulators will not be fully realized. The traditional concerns that industry regulators and competition authorities have had with the most significant problems in electricity markets today—collusive, predatory, and entry-deterring behavior, and the effects of mergers and changes in market structure on market power—drive the search for auction designs that achieve one general goal: improve market efficiency. A poor design (one with rules that are incomplete, inconsistent, or full of loopholes) may result in prices of energy, ancillary, and transmission services that do not reflect sellers' marginal costs or buyers' marginal willingness to pay, something prices would normally represent if they were set by an efficient market. Under a poor design, consumers will pay more for electricity and transmission services and consume less of each than they would under a design that induces an efficient market outcome. By the same token, suppliers will produce less energy and profit more than if the auctions were designed to enhance competition. In the short run, electricity prices will be higher than competitive prices, and in the long run, the allocation of resources to the generation and transmission of electricity will deviate from the socially optimal levels, if the power market auctions are designed poorly.

While auctions are nothing more than a set of rules describing how buyers and sellers will interact to determine the price and allocation of the good or commodity, understanding auction rules is central to understanding how aggressively parties will bid, who will win, and how cheaply a contract will be procured (or, in a sale auction, how much revenue will be raised through sale of the product). Auction rules (also known as activity rules or protocols) determine how the auction proceeds: who may participate, whether bids must exceed a particular threshold to be viable, whether changes to bids will be allowed, how large a bid increment must be, when bidding stops, and so forth.

As an example, one rule in the day-ahead energy auction administered by the PJM ISO details that sellers (*i.e.*, generators) may make market-based price offers to supply power that include the cost of starting the generator from an off position (referred to as startup costs), the theoretical cost of running a generator unit, were it able to run without producing electricity (called no-load costs), and the cost of supplying the next increment of energy needed when the unit is already supplying electricity to the grid. Another rule limits generator offers to less than $1000/MWh.[12] A corresponding set of rules dictate the kinds of information required from buyers submitting bids in the day-ahead energy auction.

12 See PJM Two-Settlement Business Rules, March 22, 2000, at 2.

Effective market and auction designs encourage short-term and long-term efficiency by suppressing gaming[13] or rendering it ineffective and establishing incentives for behavior by market participants' that are consistent with their objectives. For example, rules that make bids binding after a particular time in the auction, that assess penalties for scheduling changes or limit changes to particular periods, accompanied by strong monitoring and strict compliance enforcement, encourage market participants to truthfully represent their market positions, *i.e.*, reveal their true opportunity costs in their bids and generation commitment schedules.[14] An example in the PJM ISO's day-ahead energy auction is a rule that limits what generators will be paid if they are called upon by the system operator to relieve transmission congestion. In a daily wholesale energy auction, market participants may learn enough about the system to be able to accurately predict when and where the system is likely to experience congestion, and giving them the ability to predict whether particular units will become critical to relieve it, giving them greater leverage on the market price in those instances. This rule reduces the incentive some suppliers may have to misrepresent their offers in the day-ahead auction in order to influence the market price they would be paid for power they may be called upon to supply if the system becomes congested.

Auction rules enable power markets to be contested by lowering entry and transaction costs or reducing uncertainty about the probability of winning, especially for suppliers with smaller asset bases and more limited financial resources. For example, rules that lower bidding costs or impose fewer restrictions on bidding behavior increase market competitiveness. Simple and transparent rules work best to minimize the cost of participation, especially for smaller generators whose supply offers in day-ahead energy auctions helps to erode the potential market power of participants with larger shares of the generating capacity. Good rules remain neutral with respect to bilateral transactions: efficient rules neither encourage nor discourage bilateral transactions, enabling customized, long-term contracts to ensure realization of mutual advantages that are not priced explicitly in the spot or ancillary services markets.

13 Gaming is defined as a market participant's strategic behavior designed to profit from exploitation of recognized gaps or deficiencies in market structure and market rules, *e.g.*, bidding strategy in a simultaneous, multiround auction for energy and capacity that might take advantage of information revealed in early rounds about the market positions of other suppliers.

14 Opportunity cost is an economist's phrase for the value of the most desirable alternative given up when choosing an option or use of a resource. For example, choosing to buy a car may mean that no art work can be purchased for a year or more.

Part 3 ✥ Energy Market Fundamentals

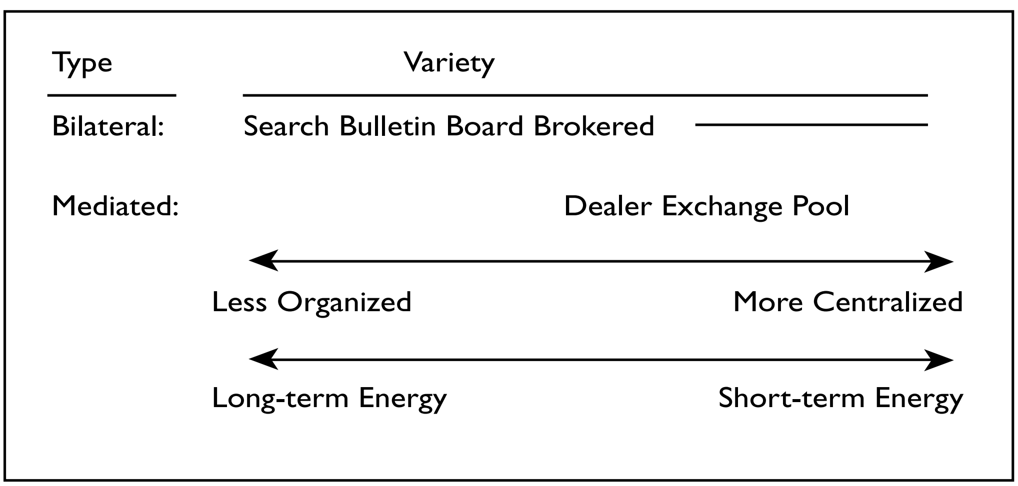
To understand the role of energy auctions and the design problems that arise with the unique features of electricity networks, energy market organization and the characteristics of electric networks warrants a brief discussion.

3.1 Organization of Energy Markets[15]

There are two basic ways to arrange trades between buyers and sellers. They can pair up—one buyer with one seller—and reach agreement on the terms of exchange (otherwise known as a bilateral trade), or sellers can sell their product to an intermediary who, in turn, sells it to final consumers. Bilateral and mediated markets come in several varieties, bilateral markets being generally less organized but there is some overlap between the types. The general spectrum of market organization is displayed in Figure 1.

Markets for particular commodities or goods often rely on a mixture of varieties from the two types. For example, the used car market is a mixture of direct search, bulletin board, and dealer markets. The New York Stock Exchange, like all exchanges, uses an auction, although when the market becomes "thin" *(i.e.*, has too few buyers and sellers), it becomes a dealer market.

Figure 1. Range of Market Types and Relation to Energy Markets



15 This section draws from Chapter 1-7, *Power System Economics* by Steven Stoft, February 2001.

Some markets work more efficiently as one type or variety and some as another. A heated debate has been waged over the years about the most appropriate form for wholesale power markets among advocates favoring more centralized forms such as exchanges or pools[16] (*i.e.*, auctions) and others who favor less organized bilateral markets such as brokered markets. Often the right answer lies with a market that makes use of both bilateral and mediated forms.

3.2 Bilateral Markets vs. Auction Markets

Compared to mediated market types, the most salient distinction of bilateral energy markets is the continual process of trading with prices unique to each transaction. Mediated energy markets (power pools and power exchanges), in contrast, are typified by a uniform price that all buyers pay and all sellers receive, with an auction run at regular times to set the market price in advance of physical delivery. The experimental and empirical evidence indicates that, in general, bilateral markets are neither more nor less competitive or efficient than exchanges or pools. Among bilateral markets with market makers, further distinctions are the "product differentiation" represented by the variety of contracts and terms tailored to individual customers and the maintenance of some degree of price continuity.

However, bilateral energy markets encounter a fundamental problem maintaining efficiency in related markets for transmission service. The demand for transmission service is derived from the interests of spatially separated buyers and sellers to reap the gains from trade between them, which can only be consummated through use of the transmission system. The maximum value of transmission service to fulfill a particular commodity contract is the sum of the gains from trade of the derived demands for the transmission system to fulfill the commodity contracts. In the aggregate, the sum of the demand values inaccurately reflects the actual demand value of transmission because it does not take into account the scarcity of the transmission capacity.

When transmission is scarce or expensive, as in the case of power transmission, market makers are challenged to use transmission facilities efficiently. They might accomplish this by aggregating transmission demands, or by brokering transmission services, but taking account of the inherent externalities in transmission of electricity, such as loop flows and thermal energy losses, there is no viable theory that assures the outcome of decentralized trading will be fully efficient. Thus, on matters of efficiency in transmission, faith in purely bilateral markets requires confidence in the ingenuity of market makers. However, this is not necessarily an argument against them, since they can operate alongside centralized exchanges,

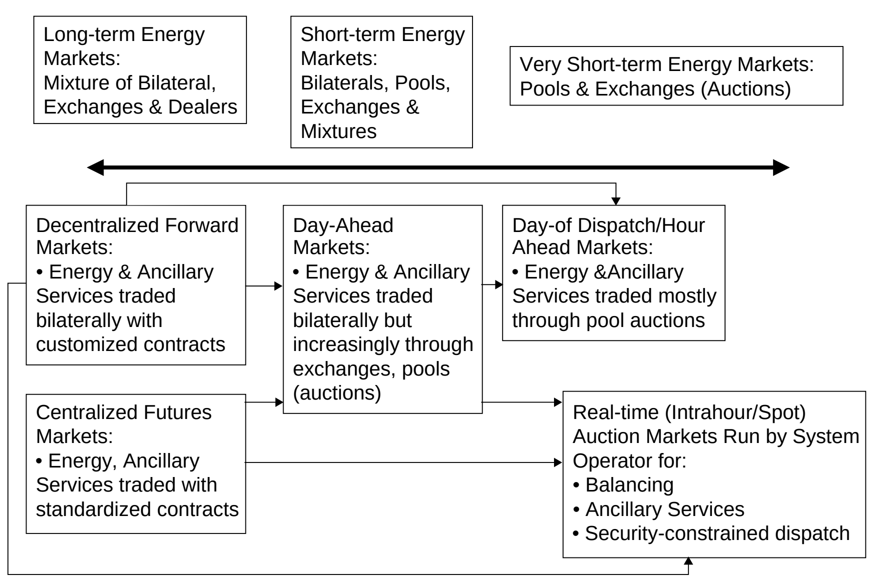
16 The term pool, derived from the term power pool, has a special meaning with regard to power markets that should not be confused with the use of the word pool when it is applied to market organization. For years, utilities in some regions have organized their production, sometimes into tight power pools that use a centralized dispatch of generators. In a deregulated market, a pool is an exchange (auction) in which the rules for supply and demand bids are often quite complex, and the system operator carries out a complex calculation to select and pay the winners.

perhaps administered by the system operator, that carry more of the responsibility at or near real-time dispatch for insuring efficient utilization of transmission facilities. For example, the California design represents a mixture of a bilateral market and a power exchange. In Scandinavia, NordPool accounts for less than 20 percent of the energy trades with the remaining 80 percent conducted through private bilateral forward contracts. Most of the wholesale electricity markets worldwide operate as a mixture of bilateral markets and power exchanges or pools.

3.3 The Relation of Market Forms to Long- and Short-term Energy

The long-term energy market, where long-term refers to future periods of a week or longer up to many years, naturally functions as a combination of both mediated and bilateral markets. (Figure 2 diagrams the relationship of market types and energy markets and time frames.) Buyers and sellers can search each other out and agree in private on the delivery of specified quantities of power at some specific times in the future through a customized contract called a *forward*. Or they can turn to a central exchange, called a *futures* market, in which standardized (regulated) contracts are traded. The transaction cost of trading in a forward market is higher than in a futures market but provides greater flexibility through customization. Buyers and sellers may seek out customized contracts that offer flexibility because so many factors that influence their predictions of electricity consumption, production and price in the future are uncertain. The flexibility enables them to make adjustments more easily and at lower cost once these factors become better known.

Figure 2. Relationships of Energy Markets and Market Forms



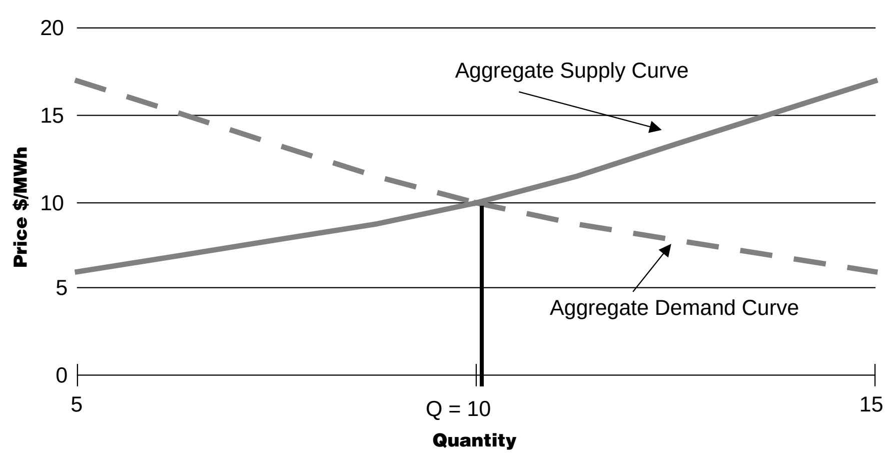
The shorter-term energy market, where shorter-term refers to a future of a day or less, down to hours before specific generators are dispatched, becomes less amenable to combinations of market types, and more amenable to mediated market forms such as exchanges and pools run as auctions. As the time to actual delivery draws near, the opportunity cost to search for an appropriate bilateral trade increases and factors affecting consumption and production become more certain, so buyers and sellers tend to find mediated markets more economical for trading, although bilateral trades are not ruled out entirely. Finally, shorter-term mediated markets enable economical exchange of energy-related products, called ancillary services, needed to support the reliable transmission of electricity. The quantities and cost of ancillary services are generally small relative to energy and, therefore, pools and exchanges relying on auctions may offer a more efficient means of exchange.

In addition, exchanges and pools (*i.e.*, auctions) are naturally efficient trading forms when the product traded is integral to the functions and services regularly performed by the transmission system operator (SO) and quantities of various products are not well known until about a day before physical trading takes place. The SO operates as an intermediary to purchase and provide ancillary services for transmission users who do not buy them directly. The SO typically relies on an auction to efficiently obtain the needed quantities in a shorter-term market, such as a day-ahead market.

3.4 Price Setting in Energy Markets: The Market-Clearing Price

Price setting in competitive energy markets adheres to the same economic principles that dictate the price outcomes in other competitive commodity or product markets. Energy auctions (pools and exchanges) can be designed to set a single uniform price or to set multiple

Figure 3. Aggregate Supply and Demand: Determination of the Market-Clearing Price



prices for a commodity.[17] Regardless of the format chosen, in a competitive energy auction, the price buyers pay and sellers receive always satisfies a fundamental economic principle: all winning buyers' valuations of the commodity are equal to or higher than the price and all winning sellers' costs of production are less than or equal to the price. The obverse of this rule is: buyers whose valuation of the commodity is lower than the price will not purchase it (these buyers are the losing buyers), and suppliers whose cost to produce the commodity is higher than the price will not be called upon to produce it (these sellers are the losing sellers).

The price that satisfies the aggregate demand of winning buyers and covers the aggregate costs of the winning sellers is called the market-clearing price, in the jargon of the electricity industry. This price clears the market in that the aggregate quantity demanded (the sum of all individual buyers' quantities demanded) and aggregate quantity supplied (the sum of all the individual sellers' quantities supplied) are equal, a condition referred to by economists as market equilibrium.[18] If the market price were set lower, there would be more quantity demanded than would be supplied, and if it were set higher, there would be less quantity demanded than supplied, in other words, the market would not be in equilibrium.

Decentralized, bilateral markets lead naturally to multiple prices determined through private contracts negotiated between pairs of buyers and sellers and dependent on private information and factors that are relevant to each party to the transaction. So, it may appear that these markets may not obey the principle. However, if all these privately negotiated prices and corresponding quantities would be revealed publicly and aggregated, the set of bilateral prices would satisfy the principle with the highest price acting as the market-clearing price.

A simple numerical example will help to illustrate the concept of a market-clearing price, (refer to Figure 3 for a stylized graphical representation of the example.) Assume three buyers and three sellers will participate in an electricity auction. Buyer 1 demands five units and is willing to pay a price of $17 per unit, Buyer 2 demands five units and is willing to pay $10 per unit, and Buyer 3 demands five units and is willing to pay $6 per unit. Seller 1 will supply five units at a price of $6 per unit, Seller 2 offers to supply five units at a price of $10 per unit and Seller 3 offers to supply five units at a price of $17 per unit. Figure 1 stylistically portrays the aggregate supply and demand curves with the price-quantity bids and offer points represented as smooth lines (the actual demand and supply curves would be step functions). In Figure 3, the aggregate demand and supply curves intersect at a quantity Q equal to ten units at a price P equal to $10 per unit. At a uniform price of $10 per unit, the quantity demanded just equals the quantity supplied, in other words, the market clears. Buyer 1 and Buyer 2 are winners paying $10 per unit for the five units they each buy, while Buyer 3, who values the five units it wants to buy at $6 per unit, loses the auction. Sellers 1 and 2 are also winners receiving $10 per unit for the five units that each is willing to sell,

17 The details of these mechanisms will be discussed in Section 4.

18 Economists often express the equality of aggregate supply and demand in terms of marginal costs and values. Thus, the market-clearing price equates the marginal cost of producing the last unit of the commodity demanded and the marginal value of the last unit of the quantity supplied.

while Seller 3 loses the auction because the costs of providing five units exceeded the market-clearing price by $7 per unit. The example demonstrates that at the market-clearing price, the value of the last unit purchased equals the cost of the last unit (or block of units) supplied. The example has been constructed assuming ideal conditions for a competitive market, sufficient numbers of buyers and sellers, access at zero cost by all to virtually the same information set, an indistinguishable commodity, and buyers and sellers who take the price as given (*i.e.*, they do not believe they can exercise any influence over the price, and, therefore, reveal in their bids and offers their true valuations and true costs). An alternative to the uniform price outcome in this example would involve paying winning suppliers their bid, referred to as a pay-as-bid (PAB) auction, discussed in Section 4 and again in Section 7.

Energy and energy-related auctions work in more or less the same fashion as the example. The auction administrator receives bids and offers from buyers and sellers, aggregates them to form aggregated demand and supply curves, finds the intersection and sets the market-clearing price and quantity on that basis. Where the auction is set up as a uniform market-clearing price, the winning buyers pay the market-clearing price and value the purchase at least as much as the price. Winning sellers receive the market-clearing price and can produce what they supply at a cost no higher than the price.

Given that all auctions will result in a market-clearing price, with winning and losing buyers and sellers, the important questions auction designers try to address are: Does market efficiency depend on the form of the auction chosen? Does the design choice affect bidder's strategies? And, if it does, how are the rules that define the auction to be laid out so that the bidder's goals are aligned with the goals of the auction?

Part 4 ✥ Auction Basics

4.1 Auction Forms and Price Setting

Auction designers have four basic types of auction formats from which to choose that have been widely analyzed and applied:

- the ascending-bid auction (also called the open, oral, or English auction),
- the descending-bid auction (used in the sale of flowers in the Netherlands and so also called the Dutch auction by economists),
- the first-price sealed-bid auction, and
- the second-price sealed-bid auction (also called the Vickrey auction by economists).

In the ascending-bid auction, the price is successively raised in rounds until one bidder remains, and that bidder wins the object at a final price set by that bidder's highest bid. This auction is characterized by the fact that bids are openly revealed, hence the frequently-used phrase open outcry, or at the very least, the best current bid is announced orally by the auctioneer or posted for all bidders to see, as in electronic auctions held on the internet.

The descending auction works in exactly the opposite way: the auctioneer starts at a very high price, and lowers the price continuously. The first bidder to declare acceptance of a price wins the object at that price.

In the first-price sealed-bid auction, each bidder independently submits a single bid, without seeing others' bids, and the object is sold to the bidder who makes the highest bid. The winner pays her bid, (*i.e.*, the price set is the highest or first price bid). In the case of electricity, the winner is paid her bid.

In the second-price sealed-bid (Vickrey) auction, each bidder independently submits a single bid without seeing others' bids, and the object is sold to the bidder with the highest bid, but the price paid is equal to the second-highest bid submitted, *i.e.*, the second price. The properties of the second-price sealed-bid auction are very similar to those of the ascending-bid auction insofar as the bidding behavior and strategy is concerned, and according to theorists, this auction type is vulnerable to collusion.

All four basic auction forms can lead to a *uniform* winning price because, in many cases, a single (unique) object or good is auctioned. However, many auctions involve the sale of multiple indistinguishable units of an object, for example, equity shares in a corporation, shares of load measured in MWh, for a given hour of the day, or shares of capacity measured in MW, to supply backup or replacement reserves in case of system contingencies for satisfying system reliability standards. In these cases, an alternative to a uniform price that does not discriminate among winning bidders, would be to discriminate by running a first-price sealed bid auction in which each winning bidder pays her bid, or is paid her offer price as in the case of an energy procurement auction run by an SO. Further discussion of the merits of uniform price vs. PAB price auctions will be taken up in Section 7.4.

Auctions can be double-sided as well as one-sided. For example, the auction is one-sided if only generators submit sealed-bid offers. If suppliers and load-serving entities, or even end-use customers, submit bids, the auction is double-sided.[19] Double-sided auctions can be run as uniform price or as sealed-bid first price (*i.e.*, PAB) auctions.

4.2 Bidding Behavior and the Revenue Equivalence Theorem

Economists in the 1980s demonstrated theoretically that the total revenue generated from any of the four basic auction forms would be equivalent under particular assumptions.[20] They showed that any auction that awards the object to the bidder with the highest private valuation and offers no expected surplus to the bidder with the lowest-feasible private valuation produces equivalent expected revenue, with each bidder making the same expected payment as a function of their private valuations. This theoretical result, called the revenue equivalence theorem, makes two extremely important statements about energy or ancillary service auctions. First, it says that the total costs to society of procuring energy or ancillary services through any of these auction forms will be the same regardless of form. Second, it says that those costs are independent of the form because buyers' and sellers' bidding strategies will be adjusted as the auction rules change.

The result applies in cases where the information that shapes buyers' and sellers' valuations of the good is strictly private (called a private-value model) and to cases where at least some of the information buyers and sellers use in forming their bids is commonly held by all, (called a common-value model). Thus, all of the standard auctions—ascending-price, descending-price, first-price, sealed-bid and second-price, sealed-bid auctions—are capable under specific circumstances of yielding the same expected revenue.

These auction forms have been studied extensively in the theoretical literature. However, this literature has been concerned primarily with the ability of these forms in one-time-only sales situations to converge to economically optimal market-clearing prices under various, relatively simple assumptions about the number and size of market participants, and the effects on their bidding strategies of aversion toward risk and the amounts and quality of information they possess. While these issues are important topics for study, they are not as important as the issues of design performance in deterring bad behavior by market participants competing in a repeated game. As Paul Klemperer has recently pointed out:

---

19 For example, suppliers could submit bids indicating at what price they would increase supply by an additional MW and consumers could submit bids indicating at what price they would be willing to decrease demand by a MW.

20 *Myerson* and *Riley* and Samuelson.

> Although the relative thinness of the auction-theoretic literature on collusion and entry deterrence may be defensible to the extent general economic principles apply, there is a real danger . . . that auction theorists will underemphasize these problems in applications. In particular, ascending, second-price, and uniform-price auction forms, while attractive in many auction theorists' models, are more vulnerable to collusive and predatory behaviour than first-price and hybrid forms. While auction theorists are justly proud of how much they can teach economics, they must not forget that the classical lessons of economics continue to apply.[21]

For energy, ancillary services and transmission rights markets, fundamental characteristics help define and classify auction formats and identify those formats that are likely to work better than others.

The most fundamental characteristic of short forward energy auctions (or an ancillary service auction) is that they are repeated daily (what economists call a repeated game). The repetition has good and bad implications for auction performance. Market participants quickly learn the rules and what bidding strategies work for them. Consequently, market-clearing prices may converge rapidly to stable distributions for particular periods (peak and off-peak) and under a variety of expected system and general conditions (*e.g.*, weather patterns). However, repetition also enables suppliers to devise bidding strategies that exploit the gaps in the rules, limits of the transmission system, or the price inelasticity of demand and supply to influence the market-clearing price. Furthermore, repetition enables all suppliers to learn about how the market works, and how the network is affected by various factors such as weather and unexpected unit outages at particular nodes. This collective knowledge can lead to the evolution of bidding and operational strategies that will have the same effect as if suppliers had agreed explicitly to share the market or fix the market-clearing price. Thus, auction formats that rely on bid revelation would be ruled out because they are likely to hasten development of these types of strategies.

The link between energy and ancillary services also shapes auction design. For any given period, a supplier can supply the energy market or ancillary services market, but not both simultaneously from the same generating unit. Therefore, bidding strategies will be shaped by expected opportunity costs and prices based on the auction settlement rule, *i.e.*, the rule that determines what price will be paid to suppliers dispatched in real time. In addition, bidding strategies will depend on whether the two sets of auctions are conducted in parallel (*i.e.*, simultaneously) or in sequence.

Thus, the principal choices for the energy and ancillary service auctions are between an ascending or a sealed bid format and between a uniform[22] (non-discriminatory) or a PAB (discriminatory) settlement format.[23] The remainder of the auction design decisions, having

---

21 *See Klemperer* 2000.

22 The uniform price auction format is also referred to as a non-discriminatory auction because all suppliers are paid the same market-clearing price, regardless of the bid. The PAB price auction format is also referred to as a discriminatory auction because the price paid to each supplier varies, that is it discriminates, among the sellers on the basis of their bids.

23 An ascending bid format will be the term we use to refer to a bidding format in which all bids are publicly revealed, such as in the open-outcry style of auction used to sell art and antiques.

to do with activity rules, depends on the technical aspects of the provision of energy and ancillary services, the inextricable link between energy and ancillary services, and the relationship of the local market to external markets via the transmission grid.

4.3 Auction Design Principles

4.3.1 Establish Objectives

Auctions need goals for designs to make sense. For energy and ancillary services, the primary goal is a least-cost dispatch of generation (and of dispatchable load) that balances supplies and load and minimizes congestion and contingency management costs.

To ensure the objective is fulfilled, it is necessary to set in place rules that ensure the prices obtained in the auction will be honored at the time of delivery. It does little good to be promised low prices in advance only to find that these prices vanish when firms default at delivery time, or are able to adjust their schedules and prices to take advantage of changes in system conditions following the close of the auction. Therefore, bidders must provide sufficient financial security to assure that they will fulfill their commitment to provide electricity at the terms bid. Further, rules must be set in place to ensure suppliers cannot legally cheat on their commitments in one market to obtain a better price for their services in another.

4.3.2 Identify the Object

One of the virtues of an auction is that it forces the auctioneer (in this case, the system operator or power exchange) to fully describe what is being purchased or sold. In many cases this is obvious. A fish auctioneer simply displays the fish for all to see (and smell). Energy and ancillary services are more complicated, but the idea applies. The more clearly the market administrator can describe what is on the block, the more competition is fostered. Particular energy and transmission support services, such as regulation or reserves, require the market designer and administrator to define products in excruciating detail. See Sidebar 1 for an illustration from the Electric Reliability Council of Texas (ERCOT) proposed protocols for the procurement of ancillary services.

In essence, energy and ancillary service markets are auctions for load responsibility. A winner in the auction is committing to meet its share of the load at specified terms. One major complicating factor is that the load is uncertain, even with decades of experience predicting the demand for electricity in terms of energy or peak capacity. The more forward of delivery (*i.e.*, consumption) the auction takes place, the greater the degree of uncertainty. Hence, while we can be clear that load responsibility is being sold, the exact quantity being sold is determined *ex post*. This leaves room for mischief from suppliers if the rules of the auction do not spell out the limits on the adjustments suppliers can make to their schedules.

Sidebar 1. Product Definition Is a Key to Auction Performance

Good auction design pays attention to the details, including carefully defining the products the system operator must procure to maintain grid stability and reliability. An excerpt from the Texas grid operator (ERCOT) Draft Protocols, submitted to the Texas Public Utility Commission in early 2001, illustrates the point. The proposed protocols call for ERCOT to provide to or purchase from qualified scheduling entities (QSEs) four types of ancillary services, one of which is called non-spinning reserve service. The protocols define this service as follows:

Non-Spinning Reserve Service

Provided by ERCOT to QSEs: Reserves maintained by ERCOT that are deployed for the Operating Hour in response to loss-of-resource contingencies on the ERCOT System.

As provided by a QSE to ERCOT: Off-line generation resource capacity, or reserved capacity from on-line generation resources, capable of being ramped to a specified output level within 30 minutes or loads acting as a resource that are capable of being interrupted within 30 minutes and that are capable of running (or being interrupted) at a specified output level for at least one (1) hour.

ERCOT Protocols, Section 6. Ancillary Services, January 5, 2001, p.4.

4.3.3 Encourage Participation by Serious Market Participants

Any good auction design encourages participation. Whether the goal is economic efficiency, maximum revenues, or lowest costs, greater participation improves auction performance. Hence, the design should reflect, in part, the preferences of potential suppliers and load-serving entities or consumers. The ways to encourage participation by serious players include:

Reduce Bidder Participation Costs

The most direct way to encourage participation is to reduce the cost of participating. Entry fees should be modest. If deposits are required, they should be fully refundable (with interest). There are tradeoffs, however, in making entry costs low. Suppliers must be required to demonstrate their credit worthiness, and this will invariably, but correctly stand as a barrier to entry for inefficient market participants.[24] The auctioneer can further reduce costs by providing detailed information on what is being procured and information that would help suppliers evaluate its value (*i.e.*, expected profitability). One might think that these costs

24 The California Public Utilities Commission learned the hard way about this tradeoff in the implementation of the retail access program when initially the rules permitted any entity that could pay a nominal registration fee to become a qualified retail energy service provider. Several hundred entities showed up on day one, many neither credit worthy nor credibly experienced in the energy service industry. When consumers began to complain of rampant charlatanism, the commission quickly tightened the requirements and reduced the field of qualified providers to about one tenth its original size.

would be insignificant, especially for the largest bidders. However, what the market designer achieves by this is to get the marginal bidder—the one that is sitting on the fence—to participate. These marginal bidders are often smaller, but nonetheless essential to vigorous competition. Without them, the large suppliers have a much greater opportunity to exercise their market power in the auction.

Limit Complexity

The complexities of the energy and ancillary services markets can easily lead to complex auction designs intended to counter any attempt by suppliers to game the system to increase their revenues and profits. Although new, computerized auction designs sometimes overcome important limitations of traditional auctions when confronted with technically complex markets such as energy, the complexity of these designs may raise participation costs. As stated above, these costs may also eliminate marginal suppliers who otherwise

Sidebar 2. Auction/Market Design Balances Complexity and Incentives

The California wholesale power market was less efficient because of the interplay between the overall market design and the ISO-run auction to procure generation and load resources to relieve intrazonal transmission congestion. Suppliers realized they could profit from flaws in the design by creating congestion in a day-ahead market that they then would be paid to relieve in the real-time market. To keep the market design simple, California's transmission grid was partitioned into only two trading zones. Grid users paid transmission prices between the zones that reflected the marginal cost of congestion relief, but did not face similar upfront costs when transmission became congested within a zone because those costs were socialized in the ISO's general operating costs averaged across all users after the fact.

At times, when the ISO determined an unconstrained intrazonal market-clearing price in the day-ahead market, the corresponding power flows implied that an intrazonal transmission path would be congested. To solve this problem, the ISO ran an auction to solicit decremental and incremental adjustment bids (incs and decs) for backing down scheduled transactions across the intrazonal congested interface or deploying replacement reserves to produce counterflows across that interface. The ISO was compelled by the rules of the auction to pay prices to suppliers on the basis of their bids, an auction design known as PAB, see Section 7.4 for further discussion.

Incs inside the constraint and decs outside the constraint are always applied in pairs to maintain system balance. The difference between the costs of the chosen inc-dec adjustment pairs determined the market-clearing price for intrazonal congestion relief paid to all resources deployed. Entities responsible for the congestion did not pay for the relief directly. Consequently, market participants were induced to over-schedule fictitious transactions across the congestion-prone intrazonal interface and collect congestion relief revenues for backing off such schedules. In a notorious case in 1999, some generators sold fictitious energy in the day-ahead market creating congestion and were paid $250/MWh in the real-time market for not injecting that energy, thus collecting substantial profit on energy that was never produced—an episode that eventually led the California ISO to establish a third trading zone. The nub of the design problem lies with over-simplification of trading zones for direct congestion pricing, leaving the ISO to pay for intrazonal congestion relief and socialize the cost through an uplift charge to all users, rather than charge grid users directly for the intrazonal congestion they caused.

might increase market contestability. Smaller suppliers may not have the resources to work out all the issues generated by the complexity. Hence, they may go into the auction with an incomplete analysis and hope for the best, or they may simply decide not to participate. Wise participants will often opt for the latter.

Rules that can be gamed are the most problematic, although almost any set of rules will, by its nature, invite suppliers to invest substantial resources to identify loopholes. Refer to Sidebar 2 for an example from California. The opportunities for and the exercise of gaming increase participation costs for everyone, and are especially harmful to smaller suppliers. Some amount of complexity may be needed for the auction to effectively meet its objectives, but a good design strikes the right balance between necessary complexity and participation costs.

Reduce Bidder Uncertainty

Bidders dislike uncertainty. Indeed, bidders will expend resources to try to reduce uncertainty. Thus, dealing with uncertainty, in the form of managing risk, is often a substantial source of participation costs. Auction rules can greatly influence the amount of strategic uncertainty that participants may face. Some uncertainty is unavoidable, but the designer should do what it can to reduce strategic uncertainty. Rules that expose the bidders to large gambles not only reduce auction efficiency, but also discourage participation by marginal bidders. The risks are simply not worth taking for these bidders.

The loss in participation due to supplier uncertainty is especially important in the standard offer service auction.[25] The reason being that, in this auction, nonparticipation is a viable option; there are generally no rules or laws requiring all capable suppliers to participate in the standard offer service auction. In most auctions, nonparticipation means that the supplier is out of luck—she does not sell energy or ancillary services. But in a standard offer service auction, an electricity supplier can still supply electricity to end users, just not as standard offer service. If the risks are too large, the supplier simply decides to compete with the auction winners after the auction, rather than during the auction. Indeed, competing *ex post* has the advantage that the supplier avoids the obligation to serve at specified prices and the risk that the regulator will attempt to second guess the market looking for a better deal. Instead, the supplier can participate in the *ex post* market and offer service that reflects the supplier's current opportunity costs.

Eliminate Possible Conflicts of Interest

Conflicts of interests can arise when one of the suppliers is affiliated with the market administrator, for example, a system operator that functions as a subdivision of a larger

25 Standard offer service is the provision of energy as part of a bundled package of generation, transmission and distribution services to retail consumers who continue to take all their electricity requirements from the local (incumbent) utility.

corporate entity (such as an energy holding company). In energy and ancillary service auctions, if the market designer and administrator is the distribution or transmission wires company and also owns generation assets or has a power marketer that may wish to participate in the auction, the incentives for self-dealing may be irresistible. This is a potential problem, since the system operator has access to within-round and between-round supplier and load information that is not public. To avoid this conflict of interest, if the generation division or power marketing affiliate elects to participate, the actual management of the auction should be delegated to a third party independent of the system operator and its affiliates. Thus, the independent market administrator implements and enforces the rules of the auction, and makes all decisions in cases where the rules allow discretion.

4.4 The Role of Activity Rules

Activity rules dictate the details of the bidding in each round of an auction, such as bid submission form, the timing of bids, the size of bid increments in a multi-round auction, or the manner in which a bidder can change a bid or exit the auction.

Activity rules serve two functions. First, they create pressure on bidders to bid actively, increasing the pace of the auction. Second, they increase the information available to bidders during an auction, particularly late in an auction that helps to ensure the price achieved is near the efficient equilibrium price and that scarce resources (*e.g.*, in the case of a financial or physical transmission rights auction) are allocated to those who value them most highly.

In uniform-price, sealed-bid energy and ancillary service auction designs, activity rules pertain more to how a supplier can offer its portfolio of resources to the various product markets, for example to provide voltage support, spinning reserves, or replacement reserves, if and when offers can be withdrawn, and what adjustments a supplier can make to unit schedules and specific unit commitments if it has been chosen to serve. Activity rules may also specify how the price will be determined from offers made under special circumstances, for example, in cases where a uniform market-clearing price cannot be calculated in the day-ahead market for some reason or what will happen when the transmission system becomes constrained. Refer to the Sidebar 3 for an illustration of these types of rules developed for bidding in energy and real-time balancing markets for the PJM ISO.[26]

As with all aspects of market design choices, there are tradeoffs between one feature and the next. For example, more flexible bidding rules allow suppliers to better represent

26 The activity rules for the CalPX were adapted from those developed for the FCC's spectrum license auctions, which have met with some success and emulated worldwide. The CalPX rules were first tested in laboratory experiments at Caltech with good results, but their practical performance was somewhat disappointing because of lack of close coordination with the California ISO.

Sidebar 3. Auction Activity Rules Guide Bidding Behavior and Encourage Participation by Serious Players

The PJM ISO conducts a day-ahead energy and real-time balancing market auction into which buyers and sellers can make bids and offers. The PJM ISO has constructed rules that govern the submission of and alterations to those bids and offers. A sample of the rules for the sellers listed here illustrates the parameters that shape the supplier's offers and attempt to control generator's behavior in making offers between the energy and balancing markets. Wording of these rules is critical for ensuring bidders are serious and that incentives to game the rules are mitigated.

For example, PJM ISO requires generators identified as Capacity Resources (*i.e.*, generation needed for maintaining system reliability) to submit binding offers in the day-ahead market (see Rules 1 and 2 below). This eliminates a valuable strategic option for a supplier—waiting until the dispatch day for better information on system conditions and the probable reliability value of specific units that can shape the supplier's offer price. But the rules also recognize the opportunity costs that suppliers may face in offering units at binding prices in the day-ahead market (see Rules 3 and 4 below). Therefore, one activity rule allows Capacity Resource generators not chosen (committed) for dispatch out of those bidding in the day-ahead market, to revise and resubmit offers to the real-time balancing market. And another rule permits suppliers to include in their offer prices the opportunity costs of a cold start and of operating in standby mode (*i.e.*, no-load operating cost) and of incremental costs of energy, which can factor in fuel costs and the expected price of energy in other markets.

The rules attempt to mitigate generator gaming between markets and market power during emergencies (see Rules 5 and 6 below). For example, PJM automatically uses day-ahead offer prices in the real-time balancing market when a generator is scheduled in the day-ahead energy market to eliminate incentives to manipulate day-ahead prices so as to receive higher prices in the real-time market. To force some discipline on market-based prices to reflect costs, Capacity Resources must submit a cost-based offer schedule in case they are called upon to relieve transmission constraints. Finally, if a supplier elected to use a cost-based offer schedule, it cannot switch back to a price-based schedule.

The following is a subset of the PJM business rules that apply to Market Sellers:

1. Generators that are Capacity Resources shall submit offers into the day-ahead market, unless they are unavailable due to forced, planned, or maintenance outages.
2. Generators that are Capacity Resources and are self-scheduling shall submit offer data in the event that they are called upon during emergency procedures.
3. If a Capacity Resource is not scheduled in the day-ahead market, it may revise its offer and submit into the real-time market or it may self-schedule the resource.
4. Generation offers may consist of startup, no-load, and incremental energy offers. A generation offer may not exceed $1000/MWh.
5. A generator offer that is accepted for the day-ahead market automatically carries over into the balancing market.
6. Each Capacity Resource must make available at least one cost-based schedule.

their costs of operations but also offer suppliers greater ability to exercise market power. Flexibility can be obtained in a number of ways including increasing the frequency of bidding (*e.g.*, once per hour vs. once per day), permitting unit specific or portfolio bids, increasing the number of bid increments, and the inclusion in bids of distinct running costs (*e.g.*, start-up, no-load, and incremental load costs).

Part 5 ❖ Markets for Energy, Capacity, and Reserves

5.1 The Role of the System Operator in Energy and Ancillary Services Markets

The key to efficient wholesale power markets is the design of the auction mechanisms for facilitating trades or allocating rights to resources (capacity or energy), and transmission rights. In agreement with this, *Joskow* has remarked:

> The core of electricity sector reform is the creation of reasonably competitive wholesale spot and forward markets for electric energy, capacity, a variety of operating reserve services (also referred to as ancillary services), plus free entry of new generating capacity to make sales in these unregulated power markets. As in other commodity markets, these markets play the traditional role of balancing supply and demand and allocating supplies among competing generators in the short run and provide economic signals for entry of new suppliers in the long run. . . . They are relied upon to provide generation resources, and economic signals for using these resources efficiently, that the operator of an electric power transmission network must rely on for maintaining the reliability and power quality of the network (frequency, voltage, and stability) and to manage congestion and related network constraints at the same speed at which electricity supply and demand attributes change—which is very fast. . . . *The(se) are also the most challenging resource allocation activities to mediate through market mechanisms.*[27]

All of the believable and working models for creating new competitive electricity markets recognize that there must be a single network operator responsible for controlling the physical operation of a control area, coordinating generator schedules, balancing loads and supply resources in real time, acquiring ancillary support services required to maintain reliability, and coordinating with neighboring control areas.

Despite a general consensus on the necessity of a system operator, there remains considerable disagreement regarding the scope of the system operator's involvement in the energy and ancillary markets. Under a bilateral contracts model, (*i.e.,* a decentralized market model) the system operator still must stand ready to make up for any imbalances that may occur in real time, maintain the physical integrity of the network, enforce transmission rights, manage conflicts between the exercise of rights to schedule generation and the actual capacity of the network to accommodate schedules, respond to emergencies, and run a settlements system to address imbalances between the generation actually delivered in real time and the actual real-time consumption.

Accordingly, the system operator must still play some role in the short-term management of the network, must contract with at least some generators for various network support and reliability services, and have adequate physical controls and communi-

27 *See* also *Joskow*, 1997.

cations links to respond to emergencies.[28] Consequently, the vast majority of the real-world applications of the hybrid pool/bilateral market model have the system operator administering forward and near-real-time auctions for energy or capacity or both on a daily basis. This involvement by the system operator in the market does not have to be in conflict with the notion that decentralized bilateral markets can be relied on to handle most of the energy and at least some of the ancillary service transactions. And in most of the extant hybrid markets today, there is no conflict.

Under a poolco model (*i.e.*, a centralized mediated market model), the system operator plays a more active role in the energy markets and in the management of markets to address network congestion.[29] The poolco model has the system operator mediating bid-based forward and real-time auction markets for delivery of power at different locations to determine day-ahead hourly market-clearing and real-time prices at each location (called a node or a bus) in the system. Simultaneously, when the unconstrained least cost dispatch cannot be realized as a result of network congestion, the system operator manages the congestion based on offers and bids from suppliers and buyers to incrementally adjust output or consumption (referred to as adjustment bids) submitted at different locations.[30] This is accomplished through the use of security-constrained mathematical optimization algorithms that take the portfolio of supply and demand bids for each hour and find the lowest cost allocations and associated clearing prices, that balance supply and demand at every node on the network given network operating constraints. In addition, early poolco model proposals fully integrated the scheduling and pricing of ancillary services (*e.g.*, operating reserves) with the scheduling and pricing of electric energy, applying overall optimization and opportunity cost pricing principles for these complementary, generator-provided services.

Under either model, the system operator's typical, but not necessary, involvement in a forward market for ancillary services (regarded properly as an energy options market) and in the real-time energy market can have a significant influence on market-clearing prices in forward energy and capacity markets and obviously can have a direct effect on prices determined in the ancillary services markets.[31] At a minimum, in the bilateral market model, the degree of influence will depend on the system operator's choice of the procurement auction for network support, reliability services and energy to manage emergencies, and its choice of how to price imbalance energy for settlements. For example, to price it at the

---

28 The bilateral model assumes, however, that bilateral trading in power, balanced scheduling obligations, self-supply of operating reserves through bilateral trading, and trading of physical transmission rights will do most of the resource allocation work, with the network operator playing a residual mopping up role.

29 In the most extreme case of a poolco model there would be no bilateral transactions.

30 A full discussion of adjustment bids is in the next section.

31 An alternative to this arrangement would be for control area operators or scheduling coordinators to arrange for provision of ancillary services, eliminating the need for the SO to administer a forward market in ancillary services. However, this simply means that the scheduling coordinators will be administering day-ahead auctions, rather than the system operator, and this may increase transaction costs for securing these services.

market-clearing price or to fold imbalance costs into an uplift charge allocated on a *pro rata* share basis. Further, if the system operator administers a day-ahead forward energy market, as in the poolco model, the design of the day-ahead auction and design choices for the ancillary services auctions will create a mode of competition among suppliers and buyers. Under either model, these choices may create incentives for generators and loads (or load serving entities) to behave in ways that are incompatible with the system operator's goals and the objectives of a competitive energy market model. Where the behaviors are incompatible with a competitive market model, the consequences will be inefficient prices and resource allocations, increasing the cost of the system operator providing network support services and managing congestion and imbalances.

Centralized designs, such as the poolco model, suffer manipulations by participants with market power that lowers market efficiency and raises problems because few instruments at the system operator's disposal are effective counter-measures. Pricing and settlement rules sufficient for incentive compatibility are often too complex to be practical, and are excluded in any case by prohibitions against price differentiation of the sort used in Vickrey auctions, while punitive sanctions and penalties for abuse are inefficient to the extent they depart from prices that measure the actual marginal costs of deviations in real time.

Many of the hybrid poolco-bilateral models in practice rely on mathematical optimization routines to determine the forward market-clearing prices based either on marginal costs of supply or market-based bids obtained from offer schedules submitted by suppliers (*e.g.*, PJM ISO and NY ISO). However, this optimization approach is a fiction when detailed knowledge of participants' costs and values is replaced by offers submitted in limited formats, and impaired further when models and algorithms are insufficient to include contingency planning that recognizes the value of flexible resources.

Decentralized designs, such as the bilateral model, are afflicted with incomplete and weakly synchronized markets that impair coordination and contingency planning to the extent that participants' self-scheduling does not fully internalize inter-temporal considerations.[32] Sanctions and penalties are replaced by potentially more efficient market prices payable for deviations from scheduled quantities, but market power can still distort prices as it does in the poolco model. The resulting equilibrium has no driver to minimize total system costs.

In sum, the case for a poolco design is strongest when there is vigorous competition and good optimization. The case for decentralization is strongest when tight coordination in forward markets is less important than good scheduling decisions by each participant, provided, of course, that a system operator manages the transmission system in real time and conducts a liquid real-time market. Whether the deficiencies of centralized designs used in practice are more important than incompleteness of those decentralized markets feasible in

32 Synchronization of the forward energy market managed by the CalPX and the ancillary services and the spot market run by the California ISO has been one of the principal problems in achieving an efficient energy market in California. See the discussion in Section 7.2.

practice depends on the situation (and changes as technology changes) and depends crucially on their comparative advantages in controlling abuses of market power and stimulating competition and entry. [33]

5.2 Spot Markets: Short-Run Operations

The real-time market (*i.e.*, real-time system operation) is a continuous, real-time supply and demand balancing and price determination process because electricity is a flow. The short-run, real-time market, otherwise known as the spot market, is relatively simple. In the short run, locational investment decisions have been made. Power plants, the transmission grid, and distribution lines are all in place. Customers and generators are connected and the work of buyers, sellers, brokers, and other service entities is largely complete. The only decisions that remain pertain to the management of the system to ensure delivery of power, which in the short run is truly a commodity product.[34]

The short run is a long time on the electrical scale, but short on a human scale—for example, half an hour or less. The spot market approximates continuous operation by revising prices every five or ten minutes, although imperfect metering and software limitations often require settlements based on a coarser time frame, such as half hourly or hourly, using the average price for the period.

In theory, over the half hour, the real-time market operates competitively to move real power from generators to loads. Generators have a marginal cost of generating real power from each plant, and customers have different quantities of demand depending on the price at that half hour. The collection of ordered generator costs define the generation merit order, from the least to most expensive plant. Refer to Figure 4 for an illustration of a merit order supply stack and corresponding demand curves that set the real-time market-clearing price.

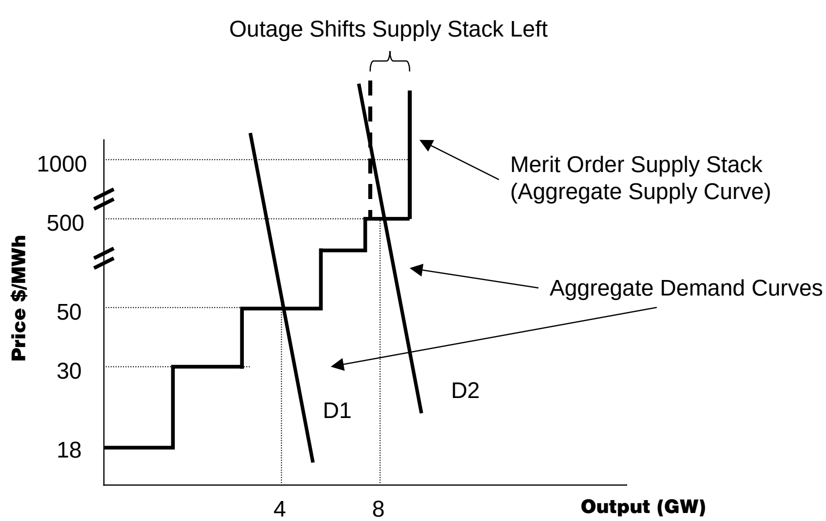
33 For further discussion of these issues, *see Wilson* 1999.

34 On the electrical scale, much can happen in half an hour and the services provided by the system include many details of dynamic frequency control and emergency response to contingencies. Due to transaction costs, if nothing else, it would be inefficient to unbundle all of these services, and many are covered as average costs in the overhead of the system. How far unbundling should go is an empirical question. For example, real power should be identified and its marginal cost recognized, but should this extend to reactive power and voltage control, or to spinning reserves required for emergency supplies? The costs associated with delivery of electric energy include many services. The direct fuel cost of generation is only one component. In analyses of energy pricing, there is no uniformity in the treatment of these ancillary services. The typical approach formulates an explicit model approximating the full electricity system, computing both a dispatch solution and associated prices for the explicit variables. Everything that is not explicit is treated as an ancillary service, for which the assumption must be that the services will be provided and charged for in some way other than through the explicit prices in the model. Given the complexity of the real electric system, such approximations or simplifications are found in every model, and there is always a boundary between the explicit variables modeled and the implicit variables that will be treated as separate ancillary services. Payment for the ancillary services may be through an average cost uplift applied to all loads.

The merit order defines the short-run marginal-cost or power supply curve whose intersection with the demand schedule for each period of time throughout a twenty-four hour period determines the market-clearing price. In Figure 4, the intersection of the aggregate supply and demand curves during a particular interval within each hour of the day sets the market-clearing price (also known as the system marginal price (SMP)). For example, when demand is four gigawatts[35] at 10 AM on a Wednesday, the market-clearing price equals $50/MWh (or 5 cents/kWh). And when demand increases, represented by a shift in the demand curve to the right (from D1 to D2), corresponding to 8 GW at 4 PM, the market-clearing price rises to $500/MWh (or 50 cents/kWh). The steepness of the demand curves reflects the fact that demand is highly inelastic in the short run (*i.e.*, in real time), nearly zero in real time. Therefore, power consumption would not change even as a result of an unanticipated generator unit outage that caused the supply curve to shift to the left, raising the market-clearing price to $1000/MWh (or $1.00/kWh).

One of the principal challenges facing auction and market designers is developing rules that account properly for intertemporal constraints on generators, such as ramp rate limits, minimum run times, and absolute upper and lower output limits.[36]

Figure 4. Intersection of Aggregate Supply and Demand Sets the Market-Clearing Price Continuously in Real Time



35 A gigawatt (GW) is a 1000 MW or a million KW.

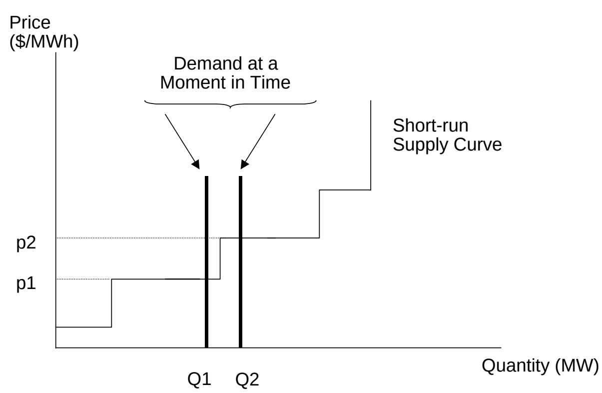
36 Eric Hirst has found these generator limits to be significant in the development of regulation markets.

Similarly, customers have demands that are price-sensitive: higher prices produce lower demands but the size and timing of the reaction depends, in part, on the size of the price increase. Market participants using bilateral transactions provide schedules with any associated bids. Generators and customers do not act unilaterally; they provide information to the system operator to be used in a decision process that determines which plants will run at any given half hour.

Because imbalances left untended can injure or destabilize transmission systems, demand and supply must be balanced continuously. Balancing is rendered more difficult by limited or expensive storage of potential energy in reservoirs, and few storage devices or backup generators at customers' sites.[37] In all designs, a system operator balances the system continuously using offers submitted in a forward spot market, usually in advance by several hours or a day, and previously acquired options for several categories of reserves.

Momentary imbalances are detected and corrected automatically by the reserve category called regulation that is provided by dispersed generators equipped with speed controls that respond to frequency sensors (called Automatic Generation Controls (AGC)).[38]

Figure 5. Merit Order Dispatch and Balancing Using Supply "Incs" and "Decs"



If demand shifts from Q1 to Q2, the increase is met from a supply "inc" at a marginal cost of p2–p1, and a market-clearing price p2. If demand shifts from Q2 to Q1, the decrease is met with a supply "dec" at the same marginal cost and clearing price.

37 This situation is changing rapidly with changes in the demand for quality power products (*e.g.*, that maintain frequency, voltage, *et cetera*).

38 Regulation does more than respond to interconnection frequency deviations. Those generators on AGC also respond to control-center signals to go up and down to maintain ACE within NERC specified bounds.

As regulation capacity nears exhaustion, its role is replaced with first preference given to offers in the spot market, followed by options on operating and replacement reserves. Operating reserves include spinning and non-spinning reserves with response times of 10 to 30 minutes. Replacement reserves (60 minutes) are activated to sustain the required margin (approximately 7%) of operating reserves as spinning and then non-spinning units are called.[39] Within each category the options are mostly used in merit order according to marginal cost as bid by the generators, but a resource option will be invoked out-of-merit order when needed to remedy an imbalance at a particular node of the transmission system when no merit-order resource will do. A further category called reliability-must-run (RMR) is contracted long term and scheduled in advance to ensure voltage support at key locations.

Each category is further divided between sources of incremental (inc) and decremental (dec) energy. The system operator meets growing demand by invoking inc supply options, and declining demand by invoking dec supply options (see Figure 5). A supply inc may be viewed as an offer to increase output at a price payable to the supplier, and a supply dec as an offer to decrease output at a price payable by the supplier; that is, a dec enables the supplier to purchase energy from the system operator to satisfy output commitments contracted previously in the forward markets that would otherwise be foregone as a result of a reduction in its own output. Thus, in a stable situation, the unused supply incs in merit order represent the extramarginal segment of the short-run supply curve at prices above the current spot price, and the unused decs represent the inframarginal segment at prices below the current spot price.[40]

In economic terms, the end result is a continually adjusted real-time price for energy, plus additional transmission costs (absorbed by the system operator, and typically passed to transmission customers through a *pro rata* charge called an uplift charge) for options exercised out-of-merit order to maintain reliability.[41] In practice, this system-wide, real-time price is defined as the highest price among those options exercised in merit order. It might seem that such a market exemplifies the economic ideal studied by theorists, but practical aspects of generator unit and transmission system operations intrude. Hirst provides an excellent discussion of the problems of managing the regulation function under generator constraints.

In most centralized pool-based systems, most of the options are not entirely voluntary and the system operator has control of real-time dispatch. Variations on this exist, however, such as in the northeastern markets, where bilateral transactions are permitted that the system operator does not have any control over. In some extreme versions of the poolco

39 A reserve unit is non-spinning if it is not synchronized with the transmission grid. Short-response, non-spinning reserves are typically provided by hydro units and combustion turbines. The regional reliability councils have differing reserve requirements, and hydro resources are allowed smaller reserve margins because they have lower forced outage frequencies.

40 Incs and decs from load or load-serving entities have the opposite interpretations.

41 An explanation for calls on generator units that would be out of the economic merit order can be found in Section 6.1.

model, a supplier will be required to bid all its operable capacity in the day-ahead market and accept assignments to reserve status. In some applications, every operating generator's incs and decs are included in the merit order even if not assigned reserve status. Further, the actual real-time dispatch is re-optimized every few minutes based on predicted demand over a rolling horizon as long as 24 hours, to take some account of ramping constraints. Typically, the economic dispatch horizon may be as short as one interval (five minutes) but more often extends out an hour or two. Thus, as difficult as it is to do, real-time (spot) market prices set in a day-ahead forward market and the prices established during the day-of dispatch and in the real-time dispatch attempt to account for anticipated inter-temporal effects.

A pure bilateral market model would operate differently. From the system operator's viewpoint, reliability would appear to be threatened because participation in forward markets for reserves and the spot market is voluntary—insufficient offers of reserves and incs and decs could jeopardize real-time operations. Forward trades on an hourly basis 24 hours in advance of or several hours before actual dispatch do not fix output rates over a given hour, so rapidly changing demand within an hour, such as the initial morning ramp up, would likely have to be met with heavy doses of regulation or other reserve products. More generally, the few categories of ancillary services for which day-ahead markets are conducted limit the system operator's flexibility, (*e.g.*, when these markets omit decremental reserves). Therefore, the system operator's insecurity about the adequacy of reserves, in part, would likely motivate it to purchase options on greater quantities of reserves than it would if it were administering a centralized (poolco) market.

The greater price volatility in bilateral markets offers a second motivation for the purchase of greater reserve quantities. In a fully decentralized system, the system operator does not control or direct dispatch except via the inc/dec options it invokes, so suppliers can deviate from day-ahead schedules, leaving the system operator responsible for balancing the system based on the generators' actual outputs (which are metered only *ex post*). The potential deviations in the real-time market from the initial schedules set the day before can be large if suppliers bypass day-ahead forward markets under an expectation of higher prices from the real-time market, or customers bypass the forward market because they expect lower real-time prices.

Single-settlement auction designs that pay generators for dispatch in real time on the basis of the market-clearing prices set in the day-ahead auctions are immune to such arbitrage, but they are vulnerable to withholding of supplies to increase the forward market spot price upon which day-ahead transactions are settled. Because arbitrage among forward and real-time markets is necessary to keep their prices linked, ideally so that the expected real-time price equals the day-ahead forward market price, the designers avoid penalizing large deviations except when they cause market failures, as in the case of onerous default prices imposed under the original design in California.

From this overview of real-time operations and spot markets it would appear that a bilateral market model is inferior. The system operator in a centralized regime can re-optimize the entire system every few minutes to redispatch all feasible resources, whereas, in a bilateral (decentralized) regime, the system operator has weaker control of a more volatile system—and both the weakness and the volatility stem from imperfections in the market structure. No decentralized system shows signs of lesser reliability, but there is evidence of higher costs for reserves, for example, in California (refer to the FERC's Order Directing Remedies for California Wholesale Electric Markets[42]).[43]

One interpretation is that contrary to appearances, the reserve markets differ in timing but not in substance. Centralized systems require participants to maintain (or acquire in secondary markets) sufficient installed and operable capacity and to bid operable capacity into the market, enabling the system operator to allocate any portion to reserves or the real-time market, whereas, in decentralized systems, the system operator conducts a daily auction to procure sufficient reserves—the end result could be the same, but evidently the decentralized implementations do not yet match the performance of the centralized procedures.

The poolco model provides a template for achieving the most efficient dispatch given the short-run marginal costs of power supply. Although dispatchable demand is not always included, there is nothing conceptually difficult about this extension.[44] The system operator controls operation of the system to achieve the efficient match of supply and demand based on the preferences of the participants as expressed through their bids. This efficient, central dispatch can be made compatible with the bilateral market outcome. The fundamental principle is that for the same load, the centralized least-cost dispatch and the competitive-market dispatch are the same. The principal difference between the poolco model and the bilateral market model solutions is the price the customer pays. In the competitive bilateral market model, the price the customer pays and the generator receives is determined privately, but disciplined by competitive choices that drive prices toward marginal costs. In the poolco model, customers pay and generators receive the average cost for any given half hour, at least on average. However, marginal cost implicitly determines the least-cost dispatch, and marginal cost is the standard determinant of competitive market pricing. Therefore, in principle, there should be little difference between the outcomes on the poolco and bilateral market models.

An important distinction between the poolco model and the bilateral market model is found in the source of the marginal-cost information for the generator supply curve. Typically, in a pure poolco model, cost data come from engineering estimates of the energy cost of generating power from a given plant at a given time. However, relying on these engineering estimates is problematic in the market model since the true opportunity costs may include other features, such as the different levels of maintenance and higher clearing prices in alternative markets, that would not be captured in the fuel cost. Replacement of the

42 See FERC Docket No. EL00-95-000 *et al.*

43 Designs for energy and ancillary services auctions in more centralized systems such as in PJM ISO, New York ISO, and ISO-NE also appear to be vulnerable to gaming by market participants that raises the costs of congestion management and real-time balancing.

44 For elaboration on the integration of the demand side into the wholesale energy and ancillary markets, *see Hirst* and Kirby.

generator's engineering estimates (that report only incremental fuel cost) with the generator's market-based bids is the natural alternative. Each bid defines the generator's reservation price that the generator would accept to run the plant in the given half hour. And these bids serve as the guide for the dispatch.

In principle, so long as generators receive the market-clearing price, and there is a sufficient number of competitors so that each generator assumes it will not be able to influence the determination of the marginal plant (and hence, the marginal cost and market-clearing price), each generator's optimal bid should be based on its short-run marginal cost. For a generator to submit a bid significantly above its short-run marginal cost would only lessen the chance of being dispatched without changing the market-clearing price. To bid less than short-run marginal cost would create the risk of being dispatched and losing money on each MWh sold.[45] A similar logic can be applied to the demand side of the equation to conclude that buyers will bid a schedule revealing their true willingness to pay.

Hence, with enough competitors and no collusion, the short-run, poolco model can elicit efficient bids from buyers and sellers. The system operator can treat these bids as the supply and demand schedules and determine the balance that maximizes benefits for producers and consumers at the market-clearing price. Hence, in the short run, electricity is a commodity, freely flowing into the transmission grid from selected generators and out of the grid to the willing customers. Every half hour, customers pay and generators receive the short-run marginal-cost price for the total quantity of energy supplied in that half hour. Everyone pays or receives the true opportunity cost in the short run. Payments follow in a simple settlement process. Such is the ideal market result that would prevail under either model when the forward and real-time markets can be synchronized under an assumption of

45 A perennial argument arises at this point against uniform market-clearing prices based on short-run marginal cost bids. As the argument goes, uniform MCPs do not permit recovery of fixed (*e.g.*, capital costs). The fact is that under a uniform market-clearing price, the only generator that does not receive payment above short-run marginal costs is the unit dispatched at the margin, that is, the unit setting the market-clearing price. All other generators dispatched by the system operator will be operating at marginal costs that are below the market-clearing price, or else they would not have been dispatched. Thus, they will receive compensation that will cover some of the capital or fixed costs of the generating unit. Generator units with short-run marginal costs below the market-clearing price are called inframarginal plants. For example, in Figure 3, the market-clearing price is \$10/MWh and the generator with marginal cost of \$6/MWh is an inframarginal plant and receives a payment of \$10/MWh minus \$6/MWh, or \$4/MWh toward recovery of its fixed costs. The supplier with marginal cost of \$10/MWh receives just that, and does not receive any revenue to cover its fixed costs. The important implication of this outcome for power markets arises in the context of transmission system reliability when very high cost units are operated very few hours per year to meet peak period demand or in emergency situations, at which times they set the market-clearing price but do not recover any fixed costs. How are these plants to be compensated to ensure that they are built and stand ready to serve in those few hours when they are needed? Capacity markets or payments may be necessary to compensate generators at the margin of the market in the absence of demand response to price rationing of scarce capacity at peak times.

sufficient numbers of buyers and sellers. Thus, the keys to success in either model are sufficient numbers of buyers and sellers and forward energy and ancillary service markets that are synchronized with the real-time dispatch market.

5.3 Forward Markets: Reserves

Poolco systems typically use suppliers' initial energy bids to assign some to reserve status, compensating those curtailed for spinning reserve the amount of their profits foregone in the energy market and paying the bids of extra-marginal units. Even so, all operating units are subject to re-dispatch in real time. Therefore, all suppliers' inc/dec bids are usually included in the merit order, even if not assigned reserve status. In contrast, participants in decentralized systems can either self-provide the required percentage of reserves or buy it from the system operator, who procures sufficient amounts of each category in a series of auctions conducted day-ahead, and additional resources contracted long-term.

The design of reserve markets has had a tortuous history that stems from three complications. The first is a legacy of the era of vertically integrated utilities with universal service obligations and simple tariffs. Because regulation of retail markets has not acknowledged the value of direct demand-side participation in energy and reserve markets, most systems make little use of reserve options on the demand side, such as contracts for service that can be curtailed by the system operator. Competition at the retail level stimulates demand-side participation in the reserve markets, but progress is slow due in part to the initial expense of installing smart meters.[46]

The second complication is that the categories of supply reserves are substitutes in a quality hierarchy based on the time a generator requires to respond to a call to provide energy. The faster response time of regulation implies that it can substitute for spinning reserve (but not the reverse), and similarly spinning reserve can substitute for non-spinning, *et cetera*. This implies that all reserve markets should be cleared simultaneously, with the result that prices should decline as generator response-times increase. These issues are discussed further in Section 5.5 on ancillary services.

The third complication is that a reserve bid has at least two dimensions, and, consequently, appropriate auction designs should have two dimensions to the price-setting process for settlements. One part is the price offered for capacity availability and the other is the price offered for energy generated when the system operator invokes its option. The theory of multi-dimensional auctions is more complicated, and judging from occasional

46 For example, in New Jersey, 60 percent of the load is hourly metered, but *the tariffs* are not hourly tariffs, thereby shielding customers from the true marginal cost of energy in real time. Other factors inhibiting progress on the retail front include a paucity of designs that permit direct demand-side participation in day-ahead and spot markets, although this is slowly being corrected.

disasters, so is practical implementation.[47] The usual fallacy is to combine the two parts by using a scoring rule (a rule that applies weights to each bid part) and accepting those bids with the lowest scores until demand is satisfied. For example, the score could be the capacity bid plus the product of the energy bid and the expected quantity of energy generated. If this expected quantity by which the energy bid is weighted is not optimally determined as a complicated function of all bids—usually it is just a constant based on the system operator's prediction of average energy requirements—then a flood of unfortunate efficiency and incentive effects can ensue.

The first effect is that the real-time energy payments do not conform to the merit order in which options must be exercised to preserve dispatch efficiency. Another effect on efficiency is that the scoring rule can attract low-cost supplies that optimally should be sold in the day-ahead energy market—this effect occurs whenever the system operator seeks to minimize the cost of its purchases rather than to maximize the gains from trade in all markets combined.[48] The incentive effects can be extreme. Each bidder recognizes that her actual chances and duration of energy generation depend on her energy bid rather than the system operator's predicted average, so she sees a tradeoff between the capacity and energy parts of her bid that encourages distorted reporting of costs. In the worst case, she thinks the system operator's prediction is wrong, for example, too high, in which case the optimal bid inflates the capacity part and deflates the energy part to zero.

Fortunately, a two-dimensional reserve auction can be reduced to a one-dimensional auction by the simple device of treating the energy bid as a reservation price and settling accounts for actual energy generation at the real-time price. That is, the scoring rule for the auction of capacity availability comprises merely the capacity bid, with zero weight given to the energy bid. The energy bid is interpreted as the real-time price below which the supplier prefers not to be called for real-time generation, so the energy bid becomes the price of the supplier's inc or dec in the merit order.

The optimality of one-dimensional auction design depends, however, on whether there are separate markets for incs and decs.[49] Creation of separate markets for incs and decs creates complications for auction design and, therefore, tradeoffs. And even though the complications have solutions, it is important to emphasize that reserve markets are widely

---

47 One notable disaster was California's 1993 Biennial Regulatory Plan Update (BRPU) auctions of power purchase agreements from Qualifying Facilities following California's implementation of the Public Utilities Regulatory Policy Act. The multi-dimensional Vickrey (*i.e.,* second price) auction selected winning bidders on the basis of a scoring algorithm (*i.e.*, weighted average of bid components) that combined energy and capacity bids. The scoring rule invited bidders to game the system by submitting extraordinarily high capacity bids and negative energy bids, tantamount to withholding of capacity to manipulate the price winners would be paid. The Federal Energy Regulatory Commission later voided the auctions.

48 This point raises an important question about what the system operator's goals should be, a question that appears to have no clear-cut answers, and, in practice, has many answers.

49 See *Chao* and Wilson.

viewed as the weakest link in decentralized designs. To some extent this is inevitable when few demand-side options are available, forcing the system operator to juggle supplies in real time to meet demands that include significant stochastic and cyclical variations. Providing the system operator with ample flexibility seems to require many markets—at least four categories of reserves that are partial substitutes, one or two of which should include decs as well as incs, and one adapted to load following. Perhaps an even better arrangement would be a unified market differentiated by a quality dimension (*e.g.*, response time) whose remuneration is determined by the system operator's opportunity cost of substituting the bid from the next slower unit. The ultimate solution, however, would involve enriching the reserve options obtained from the demand side.

5.4 Forward Markets: Energy

The variety of auction designs used in energy and reserve markets is extensive. At this point, however, it is unclear whether variety offers permanent advantages or the industry, through experimentation, will eventually converge on one design or a few designs. Evolution, but not necessarily progress, is evident in the UK's switch from a central exchange to decentralized markets for bilateral contracts.[50]

The system operator's time frame for operational control spans an hour or two, and day-ahead planning is sufficient to purchase reserves, schedule voltage support, *et cetera*. In fact, Britain's new system provides the system operator with just four hours advance notice of energy transactions. Such short horizons are possible because, in principle, a system operator accepts only balanced schedules in which energy injections equal withdrawals, so it is only in real-time operations that a system operator must cope with imbalances.[51] In most systems, however, day-ahead notice is required to provide ample time to alleviate anticipated congestion on major transmission lines. The California ISO and PJM ISO, for instance, use day-ahead markets to balance transmission on major lines so that real-time operations handle smaller local deviations.

The sequence of day-ahead followed by real-time operations for the system operator mesh with longer time frames in the energy markets.[52] For thermal generators, the basic scheduling decisions are unit commitments (startup, ramping, and running rates) made daily, so in systems with substantial thermal capacity, prices in day-ahead forward markets are basic to productive efficiency. Real-time energy demand can typically be predicted day-ahead

50 See *Wolfram* and *Sweeting*.

51 If the market model is a hybrid, as most are, the system operator will be accepting balanced schedules that include reserve services provided through bilateral arrangements. Violations of this balanced schedule principle exacerbate problems in real-time operations. Examples include failures to account for losses or for energy from units providing voltage support or reactive energy.

52 The gas industry is similar. A system operator or a pipeline company does day-ahead and intra-day scheduling while the commodity markets use long-term contracts, a monthly planning horizon, and daily scheduling.

within a few percent for each hour, and, therefore, day-ahead scheduling is usually sufficient for planning unit commitments. Longer commitments are made via bilateral contracts, some of which are physical contracts for actual production and delivery, and others, financial hedges. Within the operating day, deviations from initial schedules are common, due mainly to demand variations addressed via the spot market and by invoking options on reserves. Mature systems show a pattern of up to 80 percent contracted long term, 20 percent day-ahead, and less than 10 percent spot, although much of the supply contracted long-term actually passes through the day-ahead market. Contracts are often specified as contracts for differences (CFDs) in which the parties mutually insure each other against the difference between their contracted price and the market price.[53]

Because centralized systems consolidate all energy markets, the basic structure of the forward markets is better described in terms of a decentralized system, using California or the UK before reforms as the archetype. The topics can be divided between organizational forms and trading arrangements.

5.4.1 Organizational Forms

The two principal organizational forms for forward markets are adapted to the contracts traded. In contracts for physical delivery, the counterparty is either another market participant or the market manager.

Among those contracts between participants, nearly all are bilateral because multilateral contracts are practically infeasible. The market manager (if any) in such cases functions essentially as a broker. Some bilateral markets are merely electronic bulletin boards to which bids and offers are posted, and others offer standard contracts (*e.g.*, one is a 5/16 contract for delivery over five weekdays in the sixteen peak hours). Auxiliary terms and conditions and bundled hedges against transmission and reserve prices simulate some aspects of markets conducted by dealers, but dealer markets for pure energy are precluded by non-storability.

Those contracts in which the market manager is the counterparty are conducted as exchanges in which the manager balances aggregate demand and supply, and uses receipts from demanders to pay suppliers net of losses.

Both brokers and exchanges charge transaction fees. The contracts are termed physical because delivery is expected. However, all forward transactions are inherently financial, since commitments can be reversed by purchases or sales in the real-time market. In both forms, the typical pattern is for a participant to contract forward based on expectations but then to make adjustments based on contingencies arising the next day. A system operator's procedural rules include specific assurances that balanced energy schedules

53 Declining shares similar to those in electricity markets can be seen in fuel markets such as gas and other commodity markets, including metals, but there is an increasing tendency towards more short-term trading as electronic communication and controls improve to allow more demand-side responsiveness to spot prices.

submitted directly (from a few large participants allowed direct access to the system operator), from brokers of bilateral contracts, or from exchanges are all treated comparably, so, in principle, there is no bias in scheduling transmission or reserves.

The division of the market between long-term contracting directly or through brokers, and short-term (day-ahead or day-of) contracting through power exchanges is partly an artifact of the institutional arrangements. With few exceptions, exchanges are established at public expense as non-profit, public-benefit corporations by legislation that confines their scope to short-term markets (although a few also conduct supplementary markets for longer-term CfDs hedged against the exchange price). Their purpose is to ensure a transparent, liquid forward market whose prices can be used as less volatile benchmarks than spot-market prices. Markets for purely financial instruments, such as futures contracts, expand the influence of exchanges because they are used mainly as hedges against the exchange price and are based on the exchange's delivery points and conditions.[54]

5.4.2 Trading Arrangements

Few details are known about how bilateral contracts are privately negotiated or facilitated by brokers. In the US and Canada, several major suppliers engage in active marketing, employing traders who solicit deals and exploit arbitrage opportunities. Markets conducted via bulletin boards for posting bids and offers for standard contracts use simple trading arrangements. Similarly, markets for CfDs and price and basis swaps are conducted by telephone. The chief complication in these markets is counterparty risk—the chance that the other party to the transaction will default.[55] Public exchanges offer the obvious advantage of reduced counterparty risk.

In contrast, exchanges rely on sophisticated trading arrangements. Their authority to experiment is invariably restricted; for instance, an innovation such as a uniform second price (Vickrey) auction is precluded by prohibitions against price discrimination and a mandate to clear each hourly market independently at a uniform clearing price. But within these restrictions they have broad authority to promote efficiency. For example, bid formats are fairly rich, enabling each participant to submit a supply or demand schedule to each

54 However, Britain established one of the first day-ahead exchanges in 1989 and now plans to abolish it, relying entirely on bilateral transactions in private markets. Even though exchanges have successful records from Scandinavia to Australia, the necessity and viability of exchanges remain doubtful. California requires its power exchange to compete with bilateral markets (and also a private exchange) but arrangements in other regional power markets provide the exchange with a monopoly on short-term trades and some require bilateral contracts to pass through the exchange. If exchanges wither, their public good—in the form of a liquid, transparent market—is likely to vanish because brokered markets for bilateral contracts are intensely secretive. Efficiency could be affected because monitoring and controlling market power become difficult, and ultimately the market power of dominant brokers must be addressed.

55 A now infamous episode in 1998 convulsed the Midwestern U.S. market due to domino effects on other parties that included bankruptcies.

hourly market. Moreover, these bids are for energy only so that afterwards a supplier can conduct its own optimization of unit commitments and operating schedules. This requires generators to internalize startup costs, ramping constraints, and other considerations. Given the total energy sold in the market, this encourages productive efficiency using the supplier's private information about its costs.[56]

The Mercado in Spain offers an example of an alternative approach. The exchange rules allow a generator to withdraw tentatively accepted bids that do not meet the minimum revenue required to justify startup. Proposed designs elsewhere allow a bid format that enables a supplier to specify a minimum duration and a minimum output rate for each thermal generator, *e.g.*, the bid format originally designed for the ISO-NE. Another format enables bidders to take account of inter-temporal considerations: it uses an iterative auction so that participants can revise their bids in response to the observed pattern of prices over the 24 hourly markets for next-day delivery.[57]

The deficiencies of typical exchange procedures are obvious to economists. The bid format and market-clearing procedures take little or no account of inter-temporal factors that influence decisions of suppliers, and no contingent contracts are traded.[58] Settling trades at a uniform clearing price encourages withholding of supplies by firms with market power. An auction based on discriminatory price setting rules, such as a Vickrey auction, and most other means of strengthening incentives for generators to bid truthfully are excluded because they are discriminatory; they appear to favor generators imbued with market power by virtue of their rare flexibility. The market-clearing price is only that, it does not necessarily accurately represent the opportunity costs for suppliers and consumers that might be obtained through application of a mathematical optimization program.

The strengths of exchanges are less obvious but significant. Prices are more reflective of actual costs because suppliers schedule their plants. Settling the forward and real-time markets at their own prices suppresses gaming to affect the real-time price and optimally penalizes deviations at the real-time price. Exchanges also promote arbitrage between the forward and real-time markets, and more correctly reward flexible resources such as peaking generators. Active bidding by demanders is encouraged. In contrast, those centralized systems operating have no significant demand-side bidding programs, although work is progressing on this shortcoming. A feature of exchanges valued by participants is that prices

56 However, several exchanges operated by the system operator explicitly account for ramping constraints and start-up costs through separate side payments that may be included in the uplift costs passed on to demanders. The goal is to more closely match the operating cost curves of the generators, especially fossil-fired generators. Increasing flexibility in the bid schedules also has a down side in that it enables generators to more easily manipulate their bids to influence the market clearing price.

57 See *Wilson*.

58 A contingent contract is simply a contract whose terms of trade include consideration of contingencies, events that are uncertain at the time of the contract but whose occurrence will have material affects on the gains from trade enjoyed by the contractual parties.

are derived transparently from bids with no opaque model and arcane algorithm intervening to compute prices mathematically.

Theorists have worried most about the consequences of sequential markets. The sequence of energy markets (day-ahead, day-of, and real-time) and the parallel sequence of transmission and reserve markets, depend on rational expectations that could go awry. California appears to be an extreme case in which suppliers' and load-serving entities' expectations of better prices in the real-time and ancillary services markets, conditioned by the usage fees charged for inter-zonal transmission and by prices in reserve markets, led to significant bypass of the day-ahead forward markets.

5.5 Ancillary Services (Reserves) Markets[59]

5.5.1 Fundamental Reserve Services

At least four services would be considered in the design of an ancillary services auction: regulation, spinning reserves, supplemental (or non-spinning) reserves, and replacement reserves. Table 1, from *Hirst*, defines four real-power ancillary services that would be considered in any auction design for energy and ancillary power.

Table 1. Definitions of Four Real-power Ancillary Services.

- Regulation: generators online or on automatic generation control that can respond quickly (*i.e.*, they have a high ramp rate in MW/minute) to system operator requests for up (incremental energy) and down (decremental energy) movements.
- Spinning reserve: generators online, synchronized to the grid that can increase output immediately in response to a major outage and can reach full capacity within 10 or 15 minutes. These are often referred to as quick-start reserves.
- Supplemental (non-spinning) reserve: generators with the same capability as those satisfying spinning reserve, but that need not provide output immediately, and, therefore, the unit can be offline but capable of reaching full capacity within 10 or 15 minutes.
- Replacement reserve: generators with same capability as supplemental reserve but with a 30 or 60 minute required response time.

59 The terms reserve and ancillary will be used interchangeably in this paper to mean the set of capacity-dependent energy services that are necessary to maintain the supply and demand balance, voltage level, frequency, and system reliability.

With the exception of down regulation, all of the services are downward substitutable. A resource providing up regulation can also provide spinning reserve services, supplemental and replacement services because they are all of a lower quality than up regulation. Similarly, a generator capable of providing spinning reserve service can also provide supplemental and replacement reserve, but would not be able to provide up regulation, presumably because it is physically not capable of it.

5.5.2 Fundamental Characteristic of Reserve Markets

The characteristic feature of most reserve (ancillary services) markets is that capacity is reserved in advance (*i.e.*, usually in a day-ahead forward market), and specific generators are later ordered, or called upon, to supply incremental (or decremental) energy in response to unexpected changes in the balance of energy arising from either the supply side (*e.g.*, due to generation outages or transmission congestion) or the demand side (*e.g.*, due to unanticipated changes in load).[60]

5.5.3 Relationship of Ancillary Services Markets to Forward and Real-time Energy Markets and Prices

A generator's cost of supplying ancillary services involves direct costs and indirect (*i.e.*, opportunity) costs associated with a decision not to bid that capacity in the energy market instead. Thus, the forward market energy price and the real-time price that shapes it are determinants of the direct and indirect costs of providing ancillary services.

For generating units with variable operating costs below forward energy prices, the foregone revenue from not selling energy is an opportunity cost associated with reserving capacity in the ancillary services market. Suppliers can be expected to factor this cost into their bidding decisions in energy and ancillary markets.

Figure 6 illustrates the opportunity cost of providing ancillary services for a unit with a variable operating cost of $20, during hours when the market-clearing price for energy is $30. The generator's expected opportunity cost to reserve capacity to provide ancillary services is the difference between the day-ahead forward energy market price and the variable operating cost (*i.e.*, $10). If efficient arbitrage exists between the energy and ancillary service markets, the generator could be expected to bid capacity into the ancillary service market when clearing prices exceed $10.

For a generator with variable operating costs greater than market clearing prices in the day-ahead energy market, the direct cost of providing ancillary services is a function of the unit's operating characteristics which may include direct costs such as operation and

60 We refer here to the list of reserve services that involve generation. Ancillary services could be thought of as encompassing more than generation-based services. For example, FERC includes scheduling and dispatch functions performed by the system operator in its list of required ancillary services.

maintenance or a heat rate penalty, as well as the difference between its variable operating costs and the forward energy clearing price.

5.5.4 Reserve Market Design Options

Within a framework for achieving efficient markets, there are several formats for organizing an ancillary services auction. Each design has its own advantages and shortcomings. These alternatives have diverse efficiency and distributional implications, which have been recognized and debated in several contexts.

An efficient auction design for ancillary services that explicitly accounts for the relationship between capacity and energy markets is based on a general theory of priority pricing.[61] The distinction between the forward reservation of capacity and the real-time dispatch of energy implies that there are (or should be) two relevant prices (*i.e.*, generators' bids should have two parts). One price is set for reserved capacity and the other is set for energy. Each price is associated with a merit order, one for reserving capacity, and the other for dispatching generation in real time. These two merit orders should not be the same if the market is to be efficient.[62] Therefore, an efficient auction design for the combined energy and capacity markets requires that the two merit orders be constructed separately.[63]

---

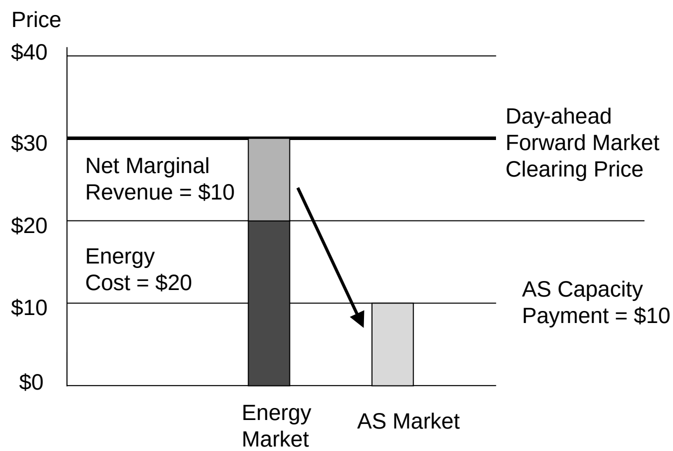
61 *See Wilson* for details on priority pricing schemes in general.

62 To appreciate this point, consider an inframarginal plant in the day-ahead energy market. This plant typically will have a higher opportunity cost to reserve capacity to serve the ancillary market than a plant that is marginal in the day-ahead energy market. This is because for the inframarginal plant to reserve capacity for the ancillary market the next day it must forego greater profits from selling that same capacity to other markets.

63 Despite the economic arguments favoring two-part bids for efficient procurement of ancillary services, the California ISO concluded in late 1999 that the single-part bidding scheme was more efficient. In the report the ISO filed with FERC, it stated:

> Recent market performance supports the conclusion that the single-part approach is more efficient and results in lower overall costs due to the significant supply of supplemental energy in the real time energy market during most hours. Since suppliers of A/S capacity must compete against this supply of supplemental energy in the real time market, units selected to provide A/S capacity have an incentive to submit competitive energy bids under the single-part approach. Since the ISO began operations, supplemental energy bids have accounted for over 70% of the energy dispatched for real time incremental energy, reflecting the fact that, for most hours, there is a deep and liquid market for real time energy. While the single-part bid approach provides incentives for bidders to bid close to their actual incremental costs, the two-part bid approach would create incentives for suppliers to modify their bidding behavior to be less reflective of actual costs. Under the two-part approach, units with a high probability of being dispatched would have an incentive to increase their energy bid prices, since they would be paid their bid price rather than the market-clearing price for imbalance energy. At the same time, units with a low probability of being dispatched could increase their capacity payment under the two-part bid approach by decreasing their energy bid price. Thus, compared to the single-part approach, the two-part approach would result in less efficient dispatch and higher overall A/S capacity and energy payments. The specific algorithm considered for the two-part bid evaluation

Figure 6. Opportunity Cost of Serving the Ancillary Services Market vs. the Energy Market



The order for reserving capacity, (*i.e.*, for determining what generators will be reserved for later call), will be based on the capacity reservation bids. For those generators selected in the capacity market, the dispatch merit order will be based on the energy price bids. Finally, the settlement price will be based on the real-time market price of energy, (*i.e.*, the most expensive unit dispatched to meet the load in real time). A supplier will only be dispatched if its bid price is no greater than the real-time price. This design has been shown theoretically to be sufficient for efficiency and incentive compatibility. It may also be a necessary feature, since to the extent that the settlement price is linked to suppliers' bid prices, there is the potential for suppliers to adapt their bidding strategies over time to exploit the connection and confound the determination of the efficient dispatch order.

The key to the efficiency of this approach is market competitiveness. For the design to induce bids (*i.e.,* reservation prices) from suppliers that reflect their marginal cost of supply, each supplier must believe she is a price taker, rather than a price setter. If she believes she can set her reserve bids to exploit contingencies or that the payment she receives is conditional on transmission conditions or on the outcomes in subsequent auctions, she will, therefore, alter her bidding strategy accordingly to attempt to influence the market clearing price.

approach is not guaranteed to minimize total capacity and energy costs. This is because this algorithm would weight each unit's energy bid by the same factor, although units are actually dispatched in merit order based on their energy bid price.

The CAISO appealed to the work of *Chao* and Wilson on multi-dimensional auctions for ancillary services. However, Chao and Wilson explicitly qualify their support for the single-part bidding scheme when they emphasize the importance of the design of the inc/dec market.

Sidebar 4. Auction Design in California Ancillary Services Market

In California, the initial design for the ancillary services market provided a separate auction for each of the four categories of ancillary services, run in sequence rather than simultaneously. In addition, the system operator was obligated to purchase the necessary quantities of each category of reserves regardless of the price set in each (*i.e.*, run as a PAB auction, see Section 7.4). Suppliers submitted two-part bids consisting of a reservation bid for capacity of the resource and an energy bid that could be used in real time for dispatching. A simplified auction evaluated the reservation bids to select reserves and determine winning bidders for a particular service, with the marginal capacity bid setting the clearing price used to pay all winners in a category. Bidders not selected could rebid (with possibly different prices) in any subsequent capacity service auction. Instances of prices increasing with generator response times, referred to as "price reversals," revealed the design problem, but not before prices for some low quality reserves were a thousand times higher than would have been expected under normal conditions. Efforts to design procedures and software to clear the four main reserve markets simultaneously, while taking account of the implicit unidirectional product substitutability provides a lesson in the practical difficulties of implementing markets for multiple goods that are capable of being both substitutes and complements, even when the theory may be clear and simple.

One auction design choice would be to procure the services through a *sequentially resolved* auction (of the type used initially in California). In this format, generators can bid a resource at a different price for each hour of the following day for each of the defined ancillary services. Given the downward substitutability of most services, the activity rules for a sequential auction must be designed carefully to ensure the auction does not fall prey to gaming that produces price reversals (see Sidebar 5 for a discussion of the problems in California's design). Some observers argue that there is no design of a sequential ancillary services auction that would not suffer from deficiencies because of generator constraints that limit the system operator's flexibility to adjust to contingencies in real time or near real time.

Price reversals are possible in a sequential reserve services auction even when generators bid their true costs. Even worse, a sequential auction can create incentives for bidders to withhold capacity in an auction for a higher quality ancillary service, such as spinning reserves, to obtain a higher price for a lower quality ancillary service, such as supplemental or replacement reserves. This bidding strategy results in capacity shortfalls in

Table 2. Data for Illustrations of Ancillary Service Procurement Auction Designs

Service Type	Demand (MW)	Bid Capacity (MW)	Bid Price ($/MW)
Regulation	500	600 100	10 15
Spinning Reserve	500	200 300	5 20

the forward market and energy shortages in the real-time (spot) market and, therefore, price spikes or greater volatility in the wholesale spot-market price.

A numerical example may help clarify the problem. Refer to Table 2 for the details of the bids used in this example. Assume the system operator is in need of procuring 500 MW of regulation and 500 MW of spinning reserves in a day-head market. The services will be procured through two sequentially run auctions. Further, assume that any capacity not selected in the first auction may be rebid in the second.

The system operator buys 500 MW of regulation at a clearing price of $10/MW. The remaining 100 MW at $10/MW and 100 MW at $15/MW are rebid in the spinning reserve auction. In the spinning reserve auction, to satisfy the 500 MW demand, the system operator dispatches generators according to their capacity bids so that it procures 200 MW (with bid

Table 3. CalPX Average Ancillary Service Volumes and Prices for May–September 2000 (Capacity Payments Only)

NP15	Regulation Up		Regulation Down		Spinning Reserve		Non-Spinning Reserve		Replacement Reserve	
	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price
May-00	152	$38.26	193	$28.71	32	$14.23	97	$7.07	33	$14.03
Jun-00	252	$126.65	245	$88.27	34	$62.67	183	$53.08	121	$69.99
Jul-00	232	$71.82	195	$55.42	87	$33.09	159	$15.48	42	$9.67
Aug-00	393	$131.74	178	$87.51	224	$65.80	102	$34.44	134	$22.51
Sep-00	336	$107.53	156	$99.99	101	$27.41	23	$10.99	45	$11.79

SP15	Regulation Up		Regulation Down		Spinning Reserve		Non-Spinning Reserve		Replacement Reserve	
	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price	Avg. Hourly Quantity	Avg. Hourly Price
May-00	336	$28.64	296	$25.26	296	$13.64	188	$7.05	91	$14.02
Jun-00	319	$120.07	233	$75.01	298	$62.52	307	$53.07	135	$69.99
Jul-00	305	$68.50	257	$44.77	263	$20.46	339	$13.60	100	$8.36
Aug-00	182	$122.67	255	$74.77	333	$60.20	319	$36.30	126	$22.01
Sep-00	149	$113.95	202	$87.90	233	$21.71	290	$13.14	46	$10.00

Source: CalPX, November 1, 2000.

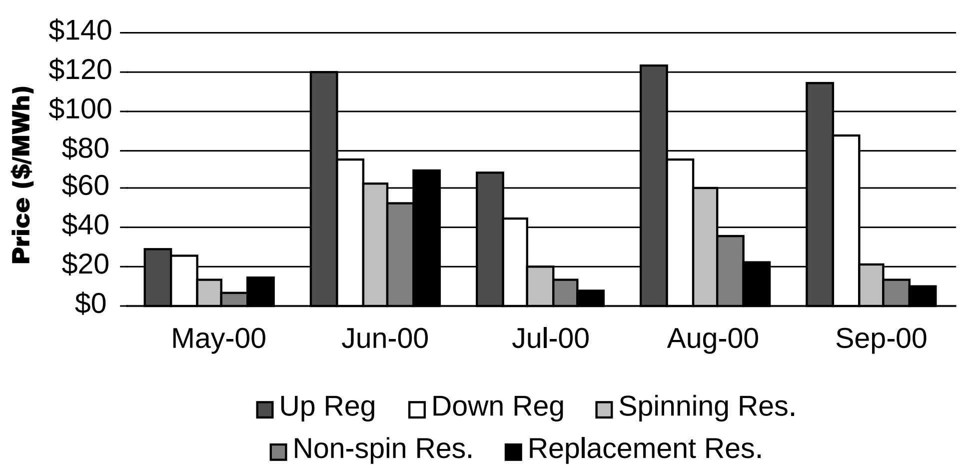
of $5/MW), 100 MW (with bid of $10/MW), 100 MW (with bid of $15/MW) and 100 MW (with bid of $20/MW). The market-clearing price for spinning reserves is set by the marginal generator at $20/MW. Thus, a price reversal occurs for these two markets with the market-clearing price for spinning reserves twice the market-clearing price for regulation, even though the bids reflect the true costs of the generators. The actual cost to provide these two services is $10,500/hr, while the actual procurement cost is $15,000/hr. With this outcome, it is easy to see how the sequential design could fall prey to stronger incentives to misrepresent available capacity or true costs, so that price reversals could be even more pronounced than this illustration.

If a sequential auction is chosen, to reduce problems of price reversals, the activity rules governing the bidding from generators and bid selection and settlement should adhere to the following principles, which apply equally as well to simultaneous formats:

- Bid monotonicity—acceptable bids should be always decreasing with service quality: lower bids must coincide with lower quality ancillary service products; and
- Market-clearing prices should be based on the maximum of all clearing prices for all services of equal or lower quality that the resource is eligible to provide.

If the bid monotonicity principle were applied consistently for ancillary service bids the result might look something like what is observed in the monthly average market-clearing prices for ancillary service reported by the California Power Exchange (CalPX) for the period May to September 2000 (see Table 3 and Figure 7). Prices reported cover capacity payments only (*i.e.,* payments for capacity reservations), for the two spatial trading zones linked by transmission path number 15 (NP15, the zone north of transmission path 15 and SP15, the

Figure 7. CalPX Avg. Ancillary Service Capacity Prices NP15 May–Sept. 2000



zone south of transmission path 15). The monthly averages may mask the fact that on certain days, price reversals may have been observed, but the relationships would be typical of what prices would be if the principle were applied consistently. Price reversals are observed in the replacement reserve market in May and June. In addition, the significant drop in the hourly replacement reserve price from June to July was the result of capping the price at $100/MWh, effective July 1, 2000. See Figure 7.

A second format, adopted by the California ISO in August 1999 and adopted in a similar format by the New York ISO and the ISO-NE in 2000, is referred to as the rational buyer approach. The design requires simultaneous bidding by generators on any or all types of ancillary services. Generators' bids must specify reserve type, capacity available, price per MW and price per MWh. The system operator's objective in selecting winning bids is to minimize the total procurement cost. The auction settlement rule pays to winning capacity bids ($/MW) of each ancillary type (as declared) a uniform market clearing price set by the highest accepted capacity bid of that type. Energy bids from suppliers are used to determine the dispatch order. Under a rational buyer approach, the system operator applies the principles stated above and adjusts purchases of individual ancillary services, so that it does not buy a low quality service when it can buy a high quality service at a lower price.[64]

To illustrate the rational buyer design, consider the previous set of bids as if they had been submitted in a simultaneous auction. Under the rational buyer approach, the system operator will accept the 600 MW bid to supply regulation, setting the market-clearing price at $10/MW. Given only 500 MW of regulation are needed, the system operator will use 100 MW of regulation at $10/MW to satisfy the spinning reserve demand, and will accept the 200 MW bid at $5/MW and the 200 MW bid at $20/MW, setting the market-clearing price for spinning reserves at $20/MW. The cost to generate these services is $11,000, $500 higher than the sequential auction approach, but the procurement cost is $14,000, $1,000 lower than under the sequential design.[65] Again, there is a price reversal in that the market-clearing price for regulation is lower than the market-clearing price for spinning reserve service.

A slight deviation on the rational buyer approach would be for the system operator to minimize operating (social) cost, *i.e.*, behave as a socially efficient buyer, buying according to the marginal cost of supply, regardless of reserve type. However, the settlement rule still sets price for a given type equal to the highest bid accepted in the category. In this case, the

64 The rational buyer approach would work even better if ancillary services procurement was further integrated with the energy market.

65 The total cost to generate the regulation services is 500 MW x $10/MW = $5000; the total cost to generate the spinning reserve services is 100 MW x $10/MW plus 200 MW $5/MW plus 200 MW x $20/MW for a total of $6000. Thus, the total cost to generate both regulation and spinning reserve services is $11,000. However, the procurement cost for regulation service is 500 MW x $10/MW = $5000 and the procurement cost for spinning reserve service is 100 MW x $10/MW plus 400 MW x $20/MWh for a total procurement cost of $9,000. Therefore, the total procurement cost for regulation and spinning reserve services under the rational buyer approach is $14,000.

system operator would take both bids in the regulation category *(i.e.*, the 600 MW bid at $10/MW, the 100 MW bid at $15/MW), using 200 MW of regulation service to satisfy the spinning reserve requirement, and would take the 200 MW bid at $5/MW and 100 MW at $20/MW in the spinning reserve category. The market-clearing price for regulation would be $15/MW and for spinning reserve $20/MW. The generation cost would be $10,500, lowest of the three approaches, while the procurement cost would be $16,500, highest of the three.

A related format is known as the smart buyer model, recommended to ISO-NE. The smart buyer format is identical to a rational buyer approach in all but one respect: the system operator reduces its demand for reserves as reserve prices increase. In conducting the markets for reserves, the system operator is effectively purchasing reliability on behalf of load. The system operator has a responsibility to purchase reserve services wisely, given the aggregate preferences of load. On this basis, the system operator, as a smart buyer, develops (*i.e.,* estimates) a demand curve for reserves that reflects the marginal value of additional reserves to the system. All generators called to supply reserve services are paid a market-clearing price determined by the intersection of the demand curve with the supply available in the system in real time. The difficult step comes in estimating a demand curve for reserves that accurately reflects the marginal value of additional reserves. There is much debate about how best to construct the demand curve for this purpose.[66]

66 For further details on the issue of constructing a demand curve to discipline ancillary service procurement costs, see *Cramton*, *DePillis* and *Chao* and DePillis.

Part 6 ❖ Constraints on Auction Designs

6.1 Supply Constraints

Energy auctions are frequently conflicted because they typically assume inter-temporal independence in production costs when setting up and running day-ahead hourly auctions. However, constraints on generators pose serious challenges to auction designers seeking to mitigate the opportunities these constraints create for suppliers to profit from strategic bidding.

Forward market auction designs must account for how generator bids (single part or multi-part) and information on unit constraints will be used by the system operator in determining unit commitments and the least-cost dispatch for each period of the following day. A paper by *Hirst* and Kirby details these constraints in the context of operating a real-time balancing market under competition and the problems that system operators (and policy makers) have in making these markets efficient and at the same time attractive for generators. They note that the complexity of the unit commitment problem and the imperfections in present-day unit commitment algorithms and computing software may enable generators to manipulate their bids to ensure particular units are included in the supply stack, perhaps at prices above competitive market-clearing levels.

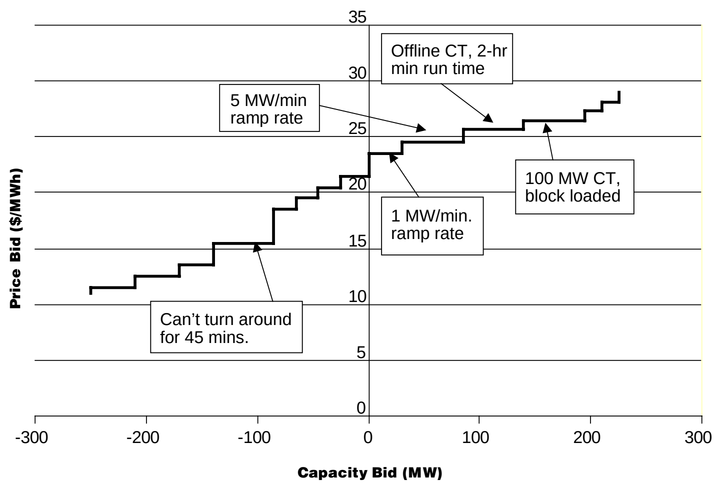
Even without bid manipulation, the constraints that limit generator unit flexibility make solving the dispatch problem difficult. If the units had no limitations, the dispatch problem would be a simple matter. However, as Hirst and Kirby point out, there are a host of constraints that include:

- minimum and maximum loadings,
- minimum and maximum run times,
- time-dependent startup costs, *i.e.*, the costs associated with bringing a unit on line from a cold condition,
- up and down ramp rates, *i.e.*, the speed (MW/minute) with which a generator can move from one output level to another, and
- block loading, *i.e.*, the situation in which a unit can be operated only at discrete (*i.e.,* block) output levels.

Figure 8 is an illustration of the kinds of limitations on generators that can influence the dispatch of units serving balancing market or other reserve service markets.

To make the task even more complicated, very often revisions to bids, or at the very least, unit schedules, are permitted after all the clearing prices are revealed, creating incentives for generators who may be pivotal suppliers at particular times of the day to adjust their schedules, that is, to change the units that will be operated in real time to serve a particular market, to take advantage of periods where clearing prices are high and supplies

Figure 8. Unit Supply Stack Based on Inc and Dec Adjustment Bids



Source: Hirst & Kirby, 2001

are inadequate to meet demand. The incentive in permitting adjustments to bids or schedules is for generators to bid a unit or a portfolio that does not truthfully represent its costs, and, thus alters the position of the unit or portfolio in the supply stack, which will be altered later once the market-clearing prices are determined. Problems of this type arose in the Victoria, Australia auction as described by *McGuire*, as well as in the California ancillary services auctions.

6.2 Demand Constraints

Complete designs for energy and some ancillary service markets would accommodate the demand as well as supply in a double-sided auction. Much debate has ensued about how best to incorporate the demand side into wholesale power markets. What can be achieved will be determined by various factors that limit voluntary participation by end-use customers or load-serving entities (LSEs) through offers to reduce load. Part of the challenge in integrating load bidding in the wholesale markets stems from the great diversity of end use technologies, usage patterns, and the value of end use appliances and technologies to consumers. The diversity would appear to be significantly greater than what is seen on the supply side.

A major constraint at the moment is technical: for load to participate directly in forward and real-time markets it must be metered appropriately and in some cases, directly controllable (*i.e.*, dispatchable) by the system operator. According to *Hirst*, "dynamic-pricing and load-reduction programs require interval meters, communication systems to move data and instructions between the customer and its LSE, and perhaps automatic-control systems that respond to time-varying prices." Hirst concludes:

> . . . all the technical components necessary for dynamic-pricing and voluntary load-reduction programs exist and have been applied in various settings. Unfortunately, the industry has not evolved to the point that standardized (off-the-shelf) equipment and communication packages are readily available. It seems that every program has to custom design its own enabling infrastructure. To the extent that complete systems involve components from various manufacturers (e.g., meters, communications systems, and data analysis software), the industry may need to develop standards to ensure that the various components can work well with each other, regardless of who manufactures what.

The challenge will be to design activity rules that permit a diverse cross-section of load to participate without creating incentives for gaming. The same principles that govern the design on the supply side apply to the issues faced on the demand side. However, designers will not be able to apply the formats for the supply side directly to the demand side. Adjustment bids from the demand side for incs and decs may be more complex or require that new or different levels of product quality be defined to accommodate constraints on load, rather than prohibiting loads from participating when they do not satisfy the criteria established for generators.

Part 7 ❖ Settlement Systems

7.1 Single Settlement Systems

In a single-settlement system, the day-ahead bids are used for scheduling but prices are determined *ex post* based on real-time dispatch. A single-settlement system consists of the following elements:

- Bids and schedules are submitted day-ahead;
- The system operator schedules units for the next day to minimize costs (using a mathematical optimization program) subject to bids, forecasts of demand, operating and transmission constraints and bilateral schedules;
- The system operator may accept bids up to an hour before real-time dispatch;
- The system operator dispatches units in real time at least cost, given the bids and forecasts for subsequent hours (*i.e.*, the system operator recomputes the optimal dispatch in real time);
- The system operator determines real-time market clearing prices as outputs from the mathematical optimization of unit dispatch;[67]
- Real-time prices are then used for all settlements to pay generators and determine charges to loads; and
- Compliance penalties (for imbalances) could be assessed against those who fail to perform as scheduled.

7.2 Multi-settlement Systems

In a multi-settlement system, the day-ahead bids are used for both scheduling and settling day-ahead transactions. Only deviations from the day-ahead schedule are priced *ex post*. The steps the system operator might follow are as follows:

- Bids and bilateral schedules are submitted day-ahead.
- The system operator schedules dispatchable units for the next day to minimize costs, given the bids, bilateral schedules, and forecast demand.

67 If the market is competitive, the interaction of supply and demand at the nodes will determine the locational prices and yield the same result as the LP optimization. For example, see *Stoft's* for an explanation of the determination of the locational market prices in a competitive market.

- The system operator determines the prices associated with the day-ahead schedule as outputs obtained from the application of the day-ahead mathematical optimization program.
- The day-ahead prices and scheduled quantities are used in the first settlement.
- The system operator may accept bid/schedule changes up to an hour before real time.
- The system operator dispatches units in real time at least cost, given the bids, schedules, and forecast demand for subsequent hours.
- The system operator determines real-time spot prices from the actual dispatch.
- Deviations from day-ahead schedules are settled at the real-time spot prices (second settlement).
- The bids submitted by suppliers in the day-ahead auction or, changes to them if allowed during adjustment periods up to the close of the auction or adjustment period, are financially binding.

A three-settlement system operates similarly, but with an hour-ahead settlement for deviations from the day-ahead schedules, and then a real-time settlement for deviations from the hour-ahead schedules.

To illustrate how the bidding in a multi-settlement system works, consider the energy market with demand-side bidding and unconstrained transmission. One day ahead, suppliers submit supply offers and demanders submit demand bids for energy for each hour and each location. Participants submitting bilateral schedules indicate the amounts to be injected and withdrawn at each location in each hour. From these bids, the system operator constructs the aggregate supply and demand curves, and identifies the market-clearing prices for each hour. Supply bids below the clearing price and demand bids above the clearing price are scheduled.

In the subsequent hour-ahead and/or real-time market, deviations from the schedule as determined by the system operator are remedied using adjustment bids; that is, incremental or decremental bids that indicate the prices at which the supplier (demander) would be willing to increase or decrease its injections (withdrawals). These incs and decs, depending on the design, may be voluntary or mandatory, and in the latter case may be deemed to be the same as the original bids.

The purpose of the incs and decs, however obtained, can be seen in an example. Suppose that a generator fails to deliver. The incs and decs are used to identify the suppliers that are best able to adjust their supplies to balance the market. In this way adjustments are made at least cost. Moreover, deviations from the day-ahead schedule are properly priced. If a generator fails to deliver, then other generators will be increased (according to their inc bids), pushing up the real-time price. The generator pays a penalty equal to the difference between the real-time price and the day-ahead price times the quantity the generator failed to deliver.

7.3 The Pros and Cons of Single- and Multi-settlement Systems

The single-settlement system may appear simpler than a multi-settlement system because it involves a single set of hourly prices and it is closer to the way tight power pools operated before wholesale restructuring. However, this simplicity is deceptive. The difficulty with the single *ex post* settlement is that much is riding on the *ex post* prices, since all earlier commitments and transactions are settled at the prices established in real time. After the day-ahead schedule is formed, bidders have an incentive to make adjustments to influence the real-time price in a favorable direction. Since the real-time price is used for all trades, the incentive for manipulation may be large.[68] For instance, day-ahead transactions, including bilateral transactions, may account for 95 percent of trades, but these are settled at prices that reflect heavily the 5 percent traded in the real-time market. Bidders can take advantage of short-term inelasticities in the supply and demand schedules to reap greater profits. Knowing how to do this is complex, and can be exploited best by large bidders with sufficient scale (*i.e.*, larger, more diverse generator portfolios) to make the efforts worthwhile. The added complexity and risk tends to discourage entry and participation by small bidders whose net revenue might be whipsawed by price volatility in the real-time market.

This gaming can be mitigated by financial penalties for failures to perform as scheduled. But this raises the question: what are the appropriate penalty values? Some flexibility is needed because of uncertainties in demand and supply. Setting the penalties too high leads to inefficient responses to this uncertainty, and setting the penalties too low leads to excessive gaming. The reliance on penalties is highly inefficient and problematic in its workings. It is a legacy of the days of tight power pools, such as NEPOOL, PJM ISO and the New York power pool, and is unworkable on a sustained basis in a competitive market. The whole idea of relying on administered penalties is inefficient, subject to endless disputes, and to continual pressure to seek modifications and exceptions.

A multi-settlement system mitigates gaming on two fronts. First, the day-ahead bids are binding financial commitments. The bids and resulting schedules are credible precisely because they are financially binding. Second, bidders are unable to alter the day-ahead prices. Day-ahead prices remain fixed for all transactions scheduled in the day-ahead market. Deviations from the day-ahead schedule affect the real-time price, but the real-time price is used only to price these deviations. Hence, in a multi-settlement system, the incentive to manipulate the real-time price is not magnified as it is in a single-settlement system. Penalties for non-performance are not needed in a multi-settlement system, since deviations from the

---

68 Contracts for differences may make the real-time price irrelevant for a significant portion of transactions. However, other bilateral transactions may be tied to the real-time price. Also, the spot price may indirectly influence the terms negotiated in bilateral contracts. The seller bids supplies into the energy market and the buyer purchases from the energy market. The seller receives the spot price; the buyer pays the spot price. However, the CfD has a strike price known only to the contract participants. If the strike price is above the spot price for the hour, the buyer pays the seller the difference; if the strike price is below the spot price, the seller pays the buyer the difference. The net payment is the contract strike price.

schedule are priced correctly. If a generator fails to deliver as scheduled, it is liable for the quantity it was supposed to deliver priced at the real-time price.

The multi-settlement system reduces risk for the bidders, since the bidders can lock in day-ahead prices. For the system operator, the multi-settlement system reduces scheduling uncertainty because it discourages schedule changes, and it automatically sets the right penalties for non-performance. The system maintains the flexibility required to respond efficiently to fluctuations in demand and supply.

One difficulty arises with multi-settlement systems because they involve multiple prices for energy. Intuition about a market-clearing price for a commodity at a particular time (and place) would lead to the belief there should be but one price. However, this is not correct. The price should be determined at the time resources are committed. Hence, if there are two commitment points (day-ahead based on forecasts and real-time based on events), then there should be two prices, one a forward price for early commitments and a second that accounts for the effects of contingencies.

Despite the advantages of multi-settlement systems, single-settlement systems can perform adequately for at least a short period of time. The initial system in the UK, and the systems in Victoria and Alberta provide examples. However, there is a strong tendency to replace them with multi-settlement systems. A multi-settlement system is now used in the UK as well as in the PJM ISO, California and New York and has been adopted in New England for implementation by early 2002. The widespread adoption of multi-settlement systems reflects the significant advantages they offer.

The differing incentive effects of the alternative settlement systems can be illustrated by a comparison of the designs in Alberta (single-settlement) and California (multi-settlement) before the FERC ordered the dissolution of the Cal PX.[69] In the Cal PX's three-part energy market, one clearing price was financially binding for trades completed in the day-ahead forward market, another clearing price was binding in the hour-ahead forward market, and the real-time price applied to ancillary services and supplemental energy purchased by the California ISO. Despite the complexity of prices, the advantage was that traders had an incentive to bid seriously in each of the forward markets since the trades concluded were financially binding at the clearing price in that market.

Alberta uses the opposite design in which all settlements are made at the final spot (*i.e.*, real-time) price, calculated *ex post*. That this design produces incentive problems can be seen in the rules required to implement it. Traders were originally prohibited from altering their day-ahead commitments, but then pressures from suppliers led to a compromise in which each trader was allowed a single re-declaration, and lately the argument has been over whether the final time for all declarations should be moved to just two hours before dispatch. These developments reflect the suppliers' strong general preferences to delay commitments until close to the time at which prices for settlement are established, so that uncertainty is reduced, and each supplier seeks to gain from committing last so that it can take maximal

69 The CalPX system was in effect until March 2001, but no longer operates.

advantage of the likely pattern of prices thereby revealed. The Alberta design also invited gaming. Importers and exporters are allowed to submit multiple virtual declarations. They have used this opportunity to declare several alternatives on a day-ahead basis and then to withdraw all but one shortly before dispatch to obtain the best terms. Of course, other traders in Alberta now demand the same privilege.

The difficulties encountered in implementing the Alberta design are intrinsic to any design in which transactions are not financially binding at the clearing price set in the market in which they are made. Having the day-ahead bids clear at the real-time price, rather than the day-ahead price, introduces a basic conflict of interest for bidders. A sequence of binding forward prices might sacrifice some efficiency in coordinating the day-ahead and real-time markets, as compared to one in which settlements are based only on real-time prices. This sacrifice may be necessary to ensure that bids made in the forward markets are serious. If viable forward markets are unnecessary, as perhaps in a pure hydro system, real-time price settlements should suffice, but it seems that justifications for forward markets also justify binding transactions at the clearing prices in these markets.[70] One must, of course, ensure that the sequence is controlled through incentive compatible activity rules.[71]

7.4 Uniform Price vs. Pay-as-Bid Settlement Systems

A debate has arisen among economists and market participants over the benefits of conducting a uniform price versus a second-price or PAB auction. [72] The arguments can be generalized into the distinctions made between non-discriminatory and discriminatory auction designs.

Advocates of PAB (discriminatory pricing) designs contend that there is significant room for tacit collusion among generators bidding into a uniform-price auction. *Armstrong*, Cowan, and Vickers first noted this potential problem was more likely to exist in power markets than other commodity markets because the interaction among the producers is repeated daily, the market clearing prices and the quantities sold by each producer are generally publicly available and the degree of uncertainty about demand is quite low in most markets. These factors facilitate an intimate knowledge of the consequences of alternative coordinated bidding strategies and allow firms to closely monitor each other, to unambiguously detect and to punish those firms that do not comply with the coordinated strategies. In

---

70 In pure hydro power supply systems, the idea of setting a forward market price is difficult when variable costs are hard to measure, and also makes it difficult to administer an inc/dec adjustment bid market.

71 By incentive compatible we mean that the rules or design for the market be consistent with the factors that motivate market participants' decisions to sell or buy in a market. If rules are not compatible with market participants' incentives, the likely result will be behavior by buyers and sellers that circumvents the rules or design to achieve their goals.

72 *See Wolfram* and *Fabra* for good examples of arguments in favor of uniform and PAB settlement formats for energy auctions.

addition, some theorists argue that uniform pricing facilitates collusion because it weakens firms' incentives to deviate from the collusive strategy.

Furthermore, PAB proponents argue that a PAB auction will lower market clearing prices because each supplier will be paid its bid, under an assumption that bids will be made at marginal cost, as they would under a competitive market model. They point out that most bids of suppliers in uniform price auctions will be lower than the market-clearing price. In a uniform price auction, its critics charge, the bids below the market-clearing price, so-called infra-marginal bids, therefore, receive payments that reward them a markup over marginal cost that over-compensates them for supplying power, and thus create incentives to bid in this fashion.

In a recent paper, commissioned by the CalPX to examine this controversy over uniform and PAB auctions, *Kahn et al* point out the fundamental fallacy of the PAB model by stating:

> The critical assumption is, of course, that after the market rules are changed, generators will bid just as they had before. *The one absolute certainty, however, is that they will not*. Knowing that unless they changed their bidding practice under the new system they would receive only their avoidable costs on their successful bids—yielding them no contribution to their fixed or common costs, let alone profits—they obviously will universally change their practice immediately, bidding instead at what they *expect* will turn out to be the market-clearing price. . . .

What Kahn and others have noted is that the same market participants who respond to the incentives of the uniform price auction would also respond appropriately to the incentives of a PAB auction, namely to bid the market-clearing price, rather than their costs. Under the revenue equivalence theorem, the results would be the same price and revenue flows as under the uniform price auction, were it not for the presence of uncertainty and transaction costs that will produce errors in the bids. The only sure thing these errors will produce is higher true costs through inefficient choices in the ultimate dispatch. There is no available evidence that the result would be lower prices, as the PAB proponents suggest. However, there are studies of both California and New England markets, where PAB rules were in effect for some markets, which suggest costs and prices would be higher.[73]

An example will elucidate some of the differences between discriminatory and uniform price auctions. Consider a day-ahead energy auction with two suppliers, Firm A and Firm B. Each supplier owns two plants. Firm A's two plants are identical, with costs of $20/MWh to generate a unit of electricity from each. Firm B has one plant with generation costs of $25/MWh and another plant with costs of $10/MWh. Assume all four plants are the same size and generate only one unit of electricity.

The auctioneer asks the firms to submit their bids in sealed envelopes without communicating with each other about their bids. Each firm submits two numbers: the price required to generate power from one plant and the price required to generate from both. For

73 Refer to *Nordhaus et al.* or *Joskow* and Kahn.

instance, if both firms submit their costs as their price offers, Firm A would bid $20 to generate from one plant and $20 to generate from both plants. Firm B would bid $10 to generate from one plant (if it was only supplying one unit of electricity instead of two, it would choose the inexpensive plant) and $25 to generate from both plants.

In a discriminatory (PAB) auction, the auctioneer buys power from the sellers who submit the lowest bids and pays each generator its bid. Under a uniform-price auction, the auctioneer buys power from the sellers who submit the lowest bids, but all successful bidders receive the market-clearing price. Assume two units of electricity will be needed. If the objective is to minimize the cost of generation, the best way of meeting that need would be to procure one unit from Firm B with cost of $10 per unit and one unit from Firm A at a cost of $20 per unit. In a PAB auction, Firm A would receive $20 and Firm B would receive $10. In a uniform price auction, the price paid to both would be $20. Note that in the uniform auction case, the auctioneer would be paying Firm B a higher price than it bid.

If each firm is well informed about what the other is likely to bid and about the number of units of electricity that will be needed, Firm B would not submit a price of $10 for the first unit of electricity. Firm B would know that it had one of the two cheapest plants available and that the competing plants all had costs of $20 and higher. Thus, Firm B would likely submit a bid slightly below $20, say $19.99. In a uniform price auction, it makes sense to submit low bids to ensure winning if the bid is unlikely to set the winning price, but this logic does not apply in a PAB auction.

Klemperer suggests that in uniform-price energy auctions, where suppliers submit offers on multiple units of electricity knowing that the price for every unit is set only by the lowest winning bid, the remainder of any firm's bidding schedule can be used as a costless threat that will determine prices only if other bidders deviate from an implicitly-agreed market division. Thus, suppliers may bid as though they have tacitly agreed to divide up the market at a very favorable price for themselves if each one bids aggressively for quantities smaller than shares obtained through explicit collusion, thus deterring other bidders from bidding for more.

Sidebar 5. Effects of Inframarginal Capacity on Bidding

Wolfram developed some evidence that inframarginal capacity had an effect on electricity bidding in England and Wales. She analyzed bids submitted by the two dominant electricity generators, National Power and PowerGen, and found three examples of the effects of inframarginal capacity. First, units with high marginal fuel costs submitted bids that reflected larger markups over their marginal costs than plants with low marginal fuel costs, consistent with the conjecture that they understand those high cost units are likely to be used after other plants are already operating and therefore, can exercise some market power to influence the price for inframarginal plants. Second, National Power submitted higher bids for similar plants, consistent with the conjecture that a block of power offered from those plants will be critical to meeting demand and set the market clearing price for all its operating units. Third, bids for a given plant were higher when more of the firm's capacity inframarginal to that plant was available.

In addition, uniform price auctions give some sellers a unilateral incentive to raise prices. For example, consider a firm with several generating plants. Each firm knows that the bids it submits for units that are likely to be marginal may set the price for all of its units. The firm has an incentive, therefore, to raise the bids for the plants likely to be marginal. If it raises the bid for a potentially marginal plant too much, that plant might not be called, and the profit from operating that plant will be lost. On the other hand, if the firm raises the bid of the marginal plant slightly and it sets the market-clearing price, the higher price is earned by all of the firm's plants. As an example of this behavior, read Sidebar 5.

In a uniform price auction, plants that earn the market-clearing price but submit lower bids are called inframarginal plants. The more inframarginal plants a firm owns, the greater the incentive it has to raise prices submitted for plants likely to be setting the market-clearing price. Similar incentives are not present in PAB auctions. In energy markets, the market-clearing price is a function of how high demand is, since with higher demand, more expensive plants need to run to meet demand. Continuing the earlier example, assume there is a distribution of possible demand levels and that bidders do not know whether there will be two units or three units of demand. If both firms knew that there would be two units of demand, their bids would lead to a market-clearing price of $20, since $20 is the price of the third most expensive plant. If both firms knew there would be three units of demand, their bids would lead to a uniform market-clearing price of $25, the price of the most expensive of the four plants. Hence, the uniform price auction may tend to drive up the price because bidders will shade their bids up on plants that are likely to be setting the market-clearing price.

While the ability to set the market-clearing price for all inframarginal plants may drive bids higher in uniform price auctions, PAB energy auctions also create incentives for bid inflation. There is a phenomenon called the winner's curse (paying more for an item or receiving less for it than the true market value) at work in markets where bidders are paid their bid and where all bidders have imperfect information about what the market-clearing price is likely to be.[74] For example, consider the two plants whose running costs are $20 each and assume now that they're owned by different firms, one each by Firm A and by Firm C. With uncertainty about whether demand would be two or three units, each firm must decide whether to bid $20 and be more likely to be called under either demand scenario but only receive $20, or take a chance and bid $25. If Firm C expected demand to be two units, perhaps because it is less optimistic than Firm A, and therefore, bid $20, it will only receive $20 when demand is three units. Knowing that this is the case, bidders try to avoid the winner's curse by submitting higher bids.

74 Winner's Curse is simply paying more for or receiving less for an item than its value, depending on whether the winner is a buyer or a seller. In the case of electricity auctions, a cursed winner is suddenly faced with the realization that other suppliers estimated a higher market value to supply the same quantity of energy. While the winning supplier has succeeded in winning the auction, the supplier has lost because profits will be lower than otherwise (a Phyrric victory). The winner will be the supplier who made the largest negative error when predicting what the market-clearing price will be.

Two other drawbacks to using a PAB auction have been raised by proponents of the uniform clearing-price model. First, the PAB auction may actually work too well. If plants that formerly submitted very low bids to ensure that they were called up under the uniform price auction end up having to guess the market-clearing price, they might sometimes guess too high and end up not selling, while more expensive plants that make better guesses do. If that happens too frequently there will be real inefficiencies in the market as plants with high marginal costs are being run before plants with low marginal costs. Consumers will pay too much if expensive plants are being run while less expensive plants sit idle. Second, by requiring suppliers to guess what the market-clearing price will be, the PAB auction raises the costs of participating for the small suppliers and may force some of them to merge in an effort to become big enough to play in the market. Thus, the PAB design may drive consolidation of the industry, increasing market concentration when the object would be to decrease concentration to heighten competition among suppliers.

If all suppliers are trying to figure out what the market-clearing price is likely to be and are bidding close to the same amount, whether or not a specific plant runs is likely to be more arbitrary and less a function of its cost. Cheap plants are likely to bid more conservatively since they stand to lose more profits if they do not get dispatched. There will no doubt be other factors that affect firms' abilities to predict the market-clearing price such as how much information the firm has about the rest of the market. As a result, there could be situations where none of the very cheap plants owned by a company that is overly optimistic about price levels will be run. Under a uniform price auction, the company could submit low bids for its very cheap plants, guaranteeing that they would be run regardless of the level of the market-clearing price.

In summary, auction theory has identified two effects on suppliers' bidding behavior. The presence of the winner's curse argues for a uniform price auction while the influence of inframarginal capacity on market-clearing prices argues for a PAB auction. In either case, the strategies of suppliers will depend in part, on how many bidders participate in the auction. In other words, strategies will depend on the degree of competition in the market and how much information each of the suppliers has relative to competitors. Whether the winner's curse or inframarginal capacity has the greater effect on the level of market-clearing prices is probably a function of specific attributes of the energy market. It is by no means clear which effect will be stronger in energy auctions. Perhaps only time will reveal the answer to this empirical question.

Part 8 ✥ Auction Designs in US Regional Power Markets

Power market auction designs have been developed to fit unique regional circumstances. Circumstances that shaped the design of the existing regional power markets have included the vestiges of vertical integration (*e.g.*, native load priority for transmission capacity), close regulatory oversight of retail market prices that insulated customers from wholesale price volatility, continued obligations for distribution utilities as providers of last resort and standard offer service, vested interests in revenue neutrality (*i.e.*, revenue preservation) and concerns about cost shifting and generator market power, especially in regions of the country subject to a transition to competition at the retail level. These issues helped shape the auction designs for particular product markets such as real-time balancing, forward energy, or ancillary services.

Designs also vary because the process that produces them is highly politicized. The process attempts to balance the disparate interests of various groups of market participants. At the wholesale level, this is partly a consequence of the insistence in Order No. 2000 on Regional Transmission Organizations (RTO) governance structures that are independent of market participants. However, market participants still have a voice in market design by their participation on stakeholder committees that offer advice to the governing boards of the regional system operators.[75] This form of governance structure, generally sanctioned by FERC, complicates the decision making for design because the outcome often reflects a compromise that makes stakeholders happy but sacrifices design features that would make the auction or the market more efficient.

8.1 Real-time Auctions

The regional power markets covered by independent system operators (*i.e.*, California, New York, New England, PJM, ERCOT) administer real-time energy and balancing markets. These are the principal markets operated by the ISOs who also administer ancillary services markets as providers of last resort. These centralized auctions (or pools) generally operate as residual markets, (*i.e.*, a majority of energy transactions are bilateral with many ancillary services also procured through bilateral contracts).[76]

---

75 As *FERC* states in Order No. 2000, " . . . even among the ISOs, there are different models of governance. As we noted in the NOPR, the dominant governance model (PJM, New England, New York, and the Midwest) for ISOs is a two-tier form of governance. The top tier consists of a non-stakeholder board, while the lower tier consists of advisory committees of stakeholders that may recommend options to the non-stakeholder board. Generally, the top tier has the final decision making authority."

76 California may be an exception to this where the California investor-owned (distribution) utilities were required to bid the majority of their load into the CalPX's day-ahead energy market. This requirement was eliminated by FERC recently.

The ISOs' real-time market auction designs are similar: bid and offer schedules are obtained in a day-ahead, double-sided auction used for scheduling and determining unit commitments for the next day. For balancing, bids and offers are submitted as incremental and decremental adjustments to scheduled quantities, with conditions on how long such adjustments can be sustained before restoration to original levels. Settlements are based on a uniform market-clearing price computed from either real-time locational marginal prices (*e.g.*, as in PJM or New York and in New England by the end of 2001) or from zonal prices (*e.g.*, as in California).[77]

Rules permitting changes to unit schedules and commitments between the close of the auction and real-time dispatch vary somewhat among these markets. On the one hand, allowing schedule changes enhances efficiency whenever such changes are consistent with an improvement in the allocation of or use of resources between buyers and sellers or whenever prohibiting a change would lead to a reduction in the benefits of exchange. On the other hand, allowing schedule changes can result in gaming by sellers who would seek naturally to profit from information they come to possess after an auction closes, especially when various day-ahead markets are organized in a cascading sequence.

Most ISOs have adopted bidding formats that generally recognize basic generator opportunity costs, but not necessarily all, such as expected prices in other markets. For example, in PJM, generators can submit two-part offer schedules that reflect energy and startup or running (*e.g.*, no-load) cost components. In ISO-NE's much anticipated multi-settlement system (MSS) generators are permitted to bid similarly but must prorate the no-load and startup costs across particular hours of the day.

Settlements in each of these markets are based on some variation of a uniform real-time hourly clearing price, computed at a system bus or averaged over a zone. In New England, the hourly settlement price is calculated *ex post* as a weighted average of real-time marginal cost prices during each hour. In New York and PJM, settlements are based on local marginal prices at nodes (*i.e.*, busses) within the region calculated as frequently as every five minutes. Table 4 summarizes the main features of the real-time balancing markets run by the ISOs.[78]

The real-time balancing markets administered by the NYISO and ISO-NE experienced problems stemming from the relative thinness of the capacity available for particular ancillary services coupled with auction settlement rules that can invite strategic behavior.

77 Zonal prices are typically computed as some weighted average of locational market-clearing (*i.e.*, marginal) prices.

78 Definitions of acronyms used in the Tables 4, 5, and 6 are: AS = ancillary services; avg = average; cap = capacity; CP = clearing price; DA = day ahead; HA = hour ahead; imb = imbalances; inc = incremental; LMP = locational marginal price; MCP = market-clearing price; mkt = market; MSS = multi-settlement system; OMO = out of merit order; op. res. = operating reserves; QSE = qualified supplier of electricity; RT = real time; sched. coor. = scheduling coordinators; TMSR = 10-minute spinning reserve; TMNSR = 10-minute non-spinning reserve; TMR = 30-minute reserve;

Consequently, the system operators occasionally had to rely on inefficient out of economic merit order dispatch to satisfy reliability criteria. In the NYISO case, the rules allowed particular generating units to set the market-clearing price that could lead to compensation for other units operated as part of a portfolio. In New England, to avoid having particular generating units set the market-clearing price, the design of the settlement system for units called out of merit order was tantamount to running a PAB auction.

For example, the NYISO allows fixed-block bidding in the real-time balancing market for units that can be scheduled. [79] Some load-serving entities, such as distribution utilities, believe this practice employs a pricing (settlement) rule not provided for in the FERC

Table 4. Real-time (RT) Energy Market

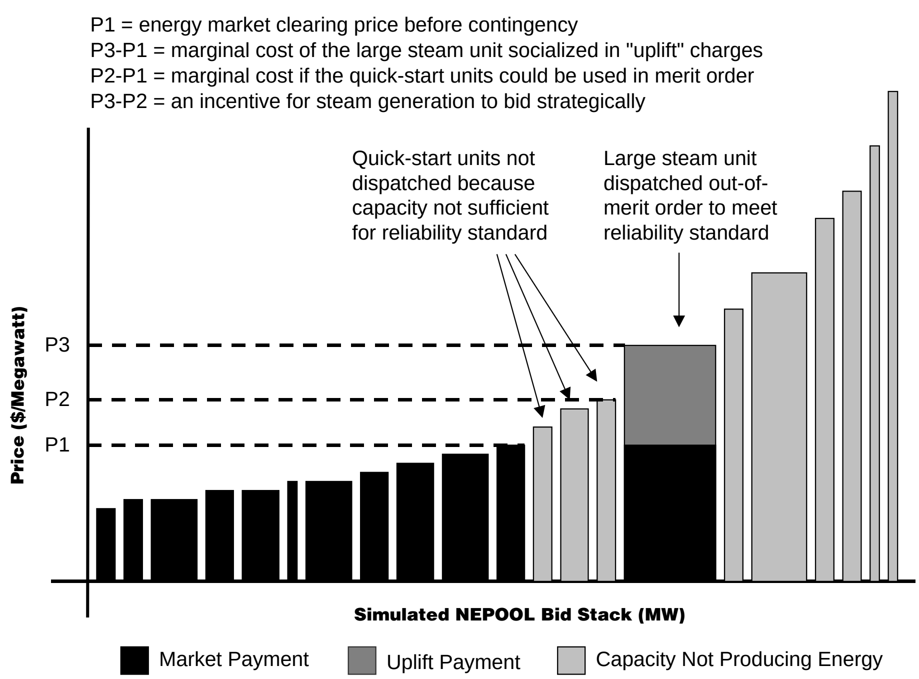
Region / Element	PJM	New York	California	New England	Texas
Settlement System	Uniform avg. hourly, LMP (RT & DA market is two-settlement system)	Uniform hourly MCP based on fixed block LMP	Uniform zonal MCP original design; now new under FERC soft cap of $150, PAB > $150 when justified	Uniform MCP set *ex post* by RT LMP, will go to new MSS early 2002	Uniform zonal MCP as set for specific ancillary services category
Bid Format	Double-sided; sealed bid; two part, submitted in DA market; market-based or cost-based bids w/ no switching after close	Double-sided; sealed bid; Single part, fixed block	Double-sided; sealed bid; single part with 11 price/quantity incrementally stepped pairs	Double-sided; sealed bid; three-part bids (hourly startup, no load costs plus 10 incremental energy prices) (new MSS) services	Separate bids for balancing up & down; balancing obtained from all other ancillary
Bidding and Schedule Flexibility	Cap. not chosen in EM can bid in RT market; self-committed units can change up to 60 mins before operating hour	Self-committed units can be changed up to 90 mins before the operating hour	Schedules and MW bids controlled and altered by sched. coords, not ISO. ISO required to maintain balanced schedules	Schedules and MW bids can be changed, but not bid prices. Units dispatched OMO to relieve congestion eligible for compensation under PAB auction format	QSEs can withdraw/modify unselected offers to RT ancillary service market through end of adjustment period

79 A fixed-block bid is an all or nothing offer from a generator to sell a large block of power (*i.e.*, MW) for a given period of time at a fixed price. The offer price, if accepted by the system operator, determines the block's order in the supply schedule, and therefore, the local market-clearing price if the block is dispatched in real time.

approved transmission tariff, (*i.e.*, that the fixed-block bid price can be used to set the locational marginal price (LMP)). When NYISO dispatches a generator with a fixed- block bid in the real-time market, other units must be backed down to accommodate it. According to New York State Electric & Gas (NYSEG), these generators are thus compelled to buy back energy at a price that exceeds their own offer price, in order to meet their contractual obligations. The NYISO compensates these displaced units for the price difference under a lost opportunity payments (LOP) provision in the ISO's tariff, a practice that NYSEG has contended is not authorized for this purpose under the tariff. Although NYSEG has not argued misbehavior on the part of generators, one consequence of this practice can be incentives for suppliers with a robust portfolio to offer fixed block bids for some units knowing that should other of their units be displaced as a result, they would be compensated under the LOP provision. In some cases, generators may know that conditions are likely to precipitate such an action by the NYISO, and thus they can increase the market-clearing price while being hedged against lost opportunities.

The initial design of the energy and reserves market in New England illustrates how market design can create perverse incentives when coupled with technical constraints on the

Figure 9. NEPOOL Design Problems



According to DePillis, NEPOOL market design application problems were caused by insufficient quick-start generator capacity and reserve price setting and settlement rules that socialized uplift costs and gave large steam generators incentives to bid up particular generators known to be needed in emergency situations.

system. Under the 1998 design, when the NEPOOL system experienced an emergency (*i.e.*, demand exceeded supply) requiring the use of quick-start plants dispatchable on a 10-minute notice, ISO-NE was forced to bypass the least-costly units in the merit order because their total capacity was insufficient to satisfy reliability reserve requirements. Instead, it dispatched reserves out-of-merit order from higher cost, coal-fired plants. This would not have been so bad except that the New England market design socialized this marginal energy cost through an uplift charge averaged across all transmission users, rather than increasing the market-clearing price that all wholesale buyers could see. Several consequences arose from this design. First, the energy market-clearing price was depressed to an inefficiently low level and, therefore, discouraged market entry by quick-start generators, so more capacity of this type was not forthcoming even though it was valued. Second, wholesale buyers did not see the true cost of energy in the price and, therefore, had no incentive to reduce their demand at the margin. Finally, suppliers' knowledge of rules on price setting and settlement for reserves dispatched out of merit order to resolve emergencies led some higher-cost generators to bid strategically to obtain uplift payments for quick start services (*i.e.*, exercise market power). Thus, they engaged in a kind of predatory pricing strategy that may have kept more efficient "quick start" units from entering the New England market. See Figure 9 for a graphical description of the problem.

8.2 Energy Markets

Basic features of the day-ahead energy markets in the regional systems are summarized in Table 5. The day-ahead energy markets in New York and California have experienced similar problems of under-scheduling by buyers in an attempt to influence prices in the day-ahead market and obtain lower prices in the real-time market. In New York, the incentive to under-schedule in the day-ahead market was a function of a rule that socialized a portion of the ISO's costs of meeting imbalances in the uplift charges allocated to all users. This rule has been changed to shift under-scheduling cost responsibilities to the buyers.

A coordination problem in California between the CalPX and the California ISO illustrates the importance of synchronizing the rules among the various markets that, if overlooked, can severely reduce market efficiency. The problem in California arose out of the separation of the day-ahead energy auction, run by the CalPX,[80] and the real-time imbalance market and the day-ahead ancillary service auctions administered by the California ISO.[81] The PX required that adjustment bids for decremental demand submitted with initial schedules for loads exceed the unconstrained energy market-clearing price. In contrast, the California ISO required adjustment bids for decremental demand to lie below the real-time

80 The existence of a separate power exchange distinguishes the California design from most other organized electricity markets. The ISO-NE, the PJM ISO, and the New York ISO operate day-ahead energy and ancillary services markets; there is no separate organized power exchange in these regional wholesale markets yet.

81 *See* the *California* ISO Special Report.

market price cap, which at one point was $500/MWh. If the unconstrained market-clearing price in the CalPX were higher than the ISO's price cap, buyers in the PX would be unable to submit adjustment bids that simultaneously satisfied both rules. This created the risk that, in the event of congestion, insufficient decremental adjustment bids would be available to relieve congestion. Thus, the CalPX requirements exposed buyers to extremely high (but constrained) congestion prices based on a default user charge equal to the California IS0's real-time price cap. Rather than take this risk, load bids were shifted out of the CalPX and into the ISO's real-time balancing market.

In addition to the inefficiencies engendered by the conflicting adjustment bid rules, the distribution utilities serving a majority of the load were restricted in their use of forward contracts and alternative markets (*e.g.*, bilateral markets) to hedge against high CalPX day-ahead prices. Consequently, they under-scheduled demand in the CalPX, shifting it into the California ISO's real-time market with the expectation of paying lower energy prices for the residual demand.[82] Suppliers correspondingly under-scheduled generation into the CalPX

Table 5. Day-ahead Energy Market

Region / Element	PJM Interconnection	New York	California	New England	Texas
Overall design; Price Cap	Voluntary, bid-based market, capped at $1000/MWh	Voluntary, bid-based market, capped at $1000/MWh	Bid-based into CalPX (closed 3-01), to be combined w/ISO	Voluntary, bid-based market w/ $1000/MWh cap; compensation for congestion re-dispatch, but can't set MCP.	Voluntary, bid-based market, no price cap
Settlement System	Single, uniform day-ahead hourly LMP	Uniform hourly LMP	Uniform hourly Zonal Avg LMP	Spring 2002, as part of the proposed MSS, will include DA LMPs, HA LMPs and RT LMPs	Uniform zonal CP
Bid Format	Double; sealed; two-part (startup & energy)	Double; sealed; two-part (startup & energy)	Double; Sealed bid; (hot/cold startup, hourly no-load, 10 inc offers)	Double; sealed bid; three-part part	Double; sealed bid; single
PX. & Reliability Functions Combined	Yes	Yes	Will be combined late 2001	Yes, centralized, after 2002	Yes
Demand Bidding	Dispatchable loads	Dispatchable loads	Dispatchable loads	Dispatchable loads	Dispatchable loads

82 The strategy for underscheduling demand amounted to bidding load into the PX day-ahead market at very low prices so that it was unlikely to clear under typical market conditions.

day-ahead market under an expectation that the California ISO's real-time market price would be higher. Consequently, the CalPX day-ahead energy market was not clearing properly.[83] The results of problems in California have been a dissolution of the CalPX day-ahead market and a shift to a price-capped real-time market that operates effectively as a PAB auction.

Aside from California, the New England markets are undergoing the greatest revision. The ISO-NE will be modifying all of its markets under a plan approved by FERC in 2000 that includes changes to the energy markets, ancillary services, and the balancing market (as well as to the congestion management system). The energy market, formerly designed as a one-part bidding and single settlement system, will initially be replaced by a three-part bidding format that is expected to do a better job of inducing generators to bid truthfully in the day-ahead hourly energy auction.

8.3 Ancillary Services Auctions

The original designs of the ancillary services procurement auctions administered by each of the regional transmission operators display considerable diversity, although there has been a convergence to a common model over the past two years. Each of the FERC-approved ISOs provides the ancillary services required under Order No. 888. All transmission customers are required to purchase energy balancing service from the ISO. Ancillary services such as spinning reserves, non-spinning reserves and replacement reserves can be self-provided and self-scheduled, obtained through bilateral contracts or obtained from the ISO. However, the ISOs are required to be providers of last resort for these services, and they generally provide them to a large percentage of their regional market's transmission customers.

Most of the ISOs' original ancillary service auction designs experienced similar problems for a variety of reasons. For example, price reversals and increasing out-of-merit-order reserve service costs were experienced in all ISO markets and often wound up in the ISO's uplift or operating costs allocated to all transmission system customers. The ability of suppliers to obtain higher prices for lower quality ancillary services was a combination of at least three things: activity rules, transmission congestion limiting supply response to higher prices, and limited or non-existent short-run demand response. In addition, some of the problems stemmed from the software design or from unrestricted sequential auctions, as in

83 Supply bid curves recently released by the PX suggest that the amount of supply being bid into the PX even at the $2,500 level during peak hours was well below the total available supply (including imports) and the level of supply needed to meet total demand. However, additional supply may have been attracted to the PX if virtually all demand was required to bid into the PX as "price takers" or at a price comparable to the potential opportunity cost of capacity in other ISO and regional markets. Another feature of centralized exchanges, one that will not materialize now that the PX has been closed, is that it lowers transaction and entry costs for smaller players and increases the likelihood of their participation in these markets.

***Table 6. Ancillary Services Markets** (refer to footnote 78 for definitions of acronyms)*

Region / Element	PJM ISO	New York	California	New England	Texas (in force June 2001)
Services	DA and HA mkts; regulation, TMSR, TMNSR, TMR, replacement	DA and HA mkts; regulation, TMSR, TMNSR, TMR	DA and HA mkts; regulation, TMNSR, replace-ment	DA and HA regulation, TMSR, TMSR, TMNSR,	Balancing down/ up, regulation, TMNSR, TMR replacement
Bid Format	Bid-based for imb. and op. res., cost-based for all others	Bid-based for all services	Bid-based, two-part capped bids, up to 7-day advanced bidding	Bid-based;	Bid-based for all when sufficient
Settlement System	Two-settlement system; DA units paid at DA MCP; HA or adjustment period units paid at HA or adjust ment period MCP	Single settlement system based on highest MCP for all eligible AS; but two-settlement system to be put in place similar to PJM ISO	Single settlement, uses MCP based on rational buyer model	MSS in place by 2002, with MCPs determined by algorithm considering overall costs of dispatch of energy and ancillary services	MSS; uniform max MCP for operating hour.
Bidding and Schedule Flexibility	Two schedules are to be submitted for all resources: a first choice and a contingency schedule	Losers can rebid in lower quality mkt; downward revision only	Considerable flexibility given participation is voluntary	Flexibility to revise bids and schedules over a 10-day period prior to the daily trading deadline prior to dispatch day	Unselected bids can be revised downward only
Demand Bidding	Dispatchable loads	Dispatchable loads	Dispatchable loads	Dispatchable loads	Dispatchable loads
Installed Capacity Mkt. (ICAP)	Twelve-month forward mkt, settled with uniform MCP; daily capacity credit mkt	Six-month forward mkt, settled w/ uniform MCP; monthly cap. deficiency auction	No capacity market created, generators ex-pected to recover all costs in uniform MCP	Monthly ICAP market operated by ISO, see discussion in Section 8.4	Capacity pay-ments for replacement by reserves when called

California. Consequently, the designs have undergone revisions over the past two years and have been converging on the rational (or smart) buyer model described in Section 4.2.4.

For example, the proposed reform to the California ancillary service markets replaces the sequential auction with a simultaneous auction in which each resource submits a single bid specifying reserve type, capacity bid and energy price if called. The ISO is allowed to substitute demand for a lower quality reserve with a higher quality and thus, use higher quality resources to meet demand for lower quality reserves. A similar scheme is applied by the NYISO and will be put in place by late 2001 by the ISO-NE, as recommended by *Cramton* and Lien.

The NYISO administers a sequential auction for ancillary services in which offers rejected for higher quality reserves can be rebid in lower quality reserve markets subject to the restriction that offer prices cannot exceed the highest market-clearing price set in preceding higher quality service auctions. This activity rule, called the rolling forward rule is the alternative approach to the simultaneous auction format for dealing with the problem of price reversals.

8.4 Capacity Markets

In each of the three ISOs in the Eastern Interconnection, an installed capacity market has been established that pays generators for reserving capacity to meet reliability requirements on the system. In these regional markets, LSEs are required to have or contract with generators for a prescribed level of reserve capacity above their peak load within a certain time frame. The specific form of the reserve requirement and the time frame over which such obligations apply varies among systems. The ISO-NE originally imposed separate requirements for installed capacity specified with respect to the annual peak and separate requirements for operable capacity specified relative to the monthly peak. It eliminated the separate requirements for operable capacity, but still maintains an installed capacity market. Formal or informal secondary capacity markets that allow trading of capacity obligations among the LSEs have accompanied reserve capacity obligations.

The reserve requirements and capacity markets provide generators with an opportunity to collect marginal revenue for their unutilized reserve generation capacity. Capacity payments also provide incentives for the building of reserves beyond those sufficient to meet short-term ancillary services needs.

The need for a capacity market and reservation payments stems largely from the fact that demand in all of these markets is highly price inelastic in the short term and does not participate directly in the short-term wholesale markets. One problem with capacity payments is that they are often set equal to the value of peaking technology capacity costs, thus generators are fully compensated even if they sit idle. Consequently, such payments may induce inefficiently high investment in capacity.

The need for a capacity payment to make up for a potential generation cost recovery shortfall can be eliminated by introducing (voluntary) load curtailment into the dispatchable supply stack as an equivalent technology with zero fixed cost and marginal cost equal to the opportunity cost to the customer (*i.e.,* the value of lost load (VOLL)). Paying generators a market-clearing price during peak periods equal to the VOLL (or more appropriately, the expected value of lost load which would be the product of the VOLL and the loss of load probability (LOLP)) when supply becomes scarce, in principle, provides them with income equivalent to a capacity payment and may mitigate the latter's over-investment incentive.

Part 9 ✥ Conclusion

If policy makers hold economic efficiency out as one of the objectives in restructuring the wholesale and retail electricity markets, then, to restate the obvious, market and auction designs matter. Designs, good or bad, will create modes of competition among buyers and sellers that will yield more or less efficient prices, resource uses, and allocations for future use. The challenge for policy makers is to provide the framework that will enable experts in design to create auctions that can induce behavior consistent with that efficiency objective.

The current federal policy that tends to endorse markets designed by committees composed of self-interested market participants should be replaced with a policy based on a sound economic framework that encourages application of design principles consistent with well functioning markets. This design by committee policy is embodied in Orders 888 and 2000, as well as in various orders approving proposed changes to the ISOs' market designs.[84] Under this policy, a design generally meets with federal approval so long as a majority of governing stakeholders vote in favor of it. This is the wrong public interest standard. The public interest threshold, for the sake of consumers, should (and must) be economic efficiency. Thus, the policy must be replaced with one based on a technical/economic framework, built upon the work of leading market designers and the experience with designs to date. Given that the rest of the country has yet to develop full-blown regional wholesale power and transmission markets under Order 2000, and all of these markets ultimately will have to be seamlessly integrated, the time is ripe for such a document to provide much needed guidance to nascent regional transmission organizations.

The policy framework should encourage incentive compatible behavior through designs that:

- Avoid socialization of generation costs in uplift charges that encourages gaming by both sellers and buyers and masks the true costs of generation at the margin in real time, sending inefficient price signals to both sides of the market;
- Closely synchronize the activity and bidding rules for short-term (*i.e.*, day-ahead) energy and ancillary service markets when they are mediated by separate agencies, such as an independent exchange for energy, and by the system operator for ancillary services;

84 The FERC has begun to provide greater guidance to the ISOs and impose stiffer standards on their proposals for redesigning their short-term energy, ancillary and transmission markets. Still, there could be greater consistency in the provision of this guidance and application of standards through a clear technical policy document.

- Examine market conditions to determine whether proposed PAB designs would be superior to uniform price designs: to see whether the winners' curse or inframarginal capacity is likely to have the greater effect;
- Introduce demand-side bidding in day-ahead energy and ancillary service auctions to discipline supply and to lead to lower reliance on or elimination of capacity markets, paying close attention to align bidding rules with the factors that motivate buyers;
- Expand on the use of multi-settlement systems and eliminate reliance on single-settlement designs to help eliminate gaming;
- Expand on the use of multi-part bid designs to better reflect opportunity costs in supplier and buyer offers and bids;
- Restrict price bids in sequential auction formats to be consistent with the rational or smart buyer model, to prevent price reversals in ancillary service markets, or better yet, rely on simultaneous auction formats for energy and ancillary services; and
- Keep in mind the revenue equivalence theorem that implies buyers and sellers will adjust their bid and offer strategies to fit the auction design so that the total cost of the auction will be equivalent regardless of the design choice.

While any one recommendation on this list cannot be applied without consideration for other elements in the overall market design, each can be a guide to building more efficient markets. The potential cost of poor (*i.e.*, less efficient) design decisions, that can be observed in the form of higher market prices, higher ancillary service costs, increased costs of out of merit order dispatch, and transfers of wealth from consumers to producers, should spur policy makers to require testing of design proposals to the extent that this can be done in advance.

The goal of energy auction design—lowest cost reliable dispatch of short-term and long-term sources of electricity available in the current market achieved through an open, transparent competitive process—can be realized in a number of ways, as this paper demonstrates. Selection of most appropriate design(s) will necessarily involve careful consideration of the individual, local market structure, the idiosyncrasies of the regional grid, and a healthy appreciation of the efficiency/fairness tradeoffs that accompany any particular design choice. A well-designed auction can encourage the open competition that is the clearest sign that the best possible terms have been found for energy and ancillary services. With firm guidance from policy makers, the industry can succeed in finding incentive compatible auction designs that will increase the economic efficiency of the power market and social welfare in the long term.

The most important issues in auction design for energy markets may well be the prevention of collusive, predatory, and entry deterring behavior. The regulatory objective should be to put in place policies that promote designs that suppress gaming or render it ineffective in favor of greater efficiency. Identifying the best locus of incentives and competitive forces becomes the central design problem and the greatest challenge for designers. The complexity of the network and its operations, the interdependency of generation and transmission products and services, and the vast number of design options ensures that market and auction designs will be less than perfect the first time. Thus, design will be an iterative process and designers may need to build in flexibility for performance review and design changes based on actual experience.

References

Akerlof, G. A., *The Market for 'Lemons': Quality Uncertainty and Market Mechanisms*, Quarterly Journal of Economics, 84, August 1970, 488–500.

Armstrong, M., Cowan, S. and Vickers, J.S., *Regulatory Reform*, Cambridge, MA: MIT Press, 1994.

Ashenfelter, O., *How Auctions Work for Wine and Art,* Journal of Economic Perspectives, Vol. 3, No. 3 (Summer 1989) 23–36.

Boyes, W. J., and S. K. Happel, *Auctions as an Allocation Mechanism in Academia: The Case of the Faculty Offices*, Journal of Economic Perspectives, Vol. 3, No. 3 (Summer 1989) 37–40.

Brough, J., *Auction!* Indianapolis and New York: Bobbs-Merrill, 1963, 26–27.

Bulow, J.I. and Klemperer, P.D., *Auctions vs. Negotiations,* American Economic Review, Vol. 86, 1996, 180–94.

Bunn, D. W. and Larsen, E. R., *Sensitivity of Reserve Margin to Factors Influencing Investment Behaviour in the Electricity Market of England and Wales,* Energy Policy, 1992, 420–429.

California ISO, Department of Market Analysis, *Report on the California Energy Market Issues and Performance: May–June 2000, Special Report*, August 10, 2000.

California Power Exchange Corporation, Compliance Unit, *Price Movements in California Power Exchange Markets, Analysis of Price Activity: May–September 2000,* November 1, 2000, Table 1, 111.

Cassady, R., *Auctions and Auctioneering*. Berkeley: University of California Press, 1967.

Chandley, J. D., Harvey, S. M., and Hogan, W.W., *Electricity Market Reform In California,* working paper, November 22, 2000.

Chao, H. and Wilson, R., *Multi-Dimensional Procurement Auctions for Power Reserves: Incentive-Compatible Evaluation and Settlement Rules, w*orking paper, Stanford

University, 1999.

Chao2, H. and DePillis, M., *A Reserve Market Reform Proposal for NEPOOL,* working paper, New England ISO, July 2000.

Chari, V.V. and Weber, R.J., *How the U.S. Treasury Should Auction Its Debt*, Federal Reserve Bank of Minneapolis, Quarterly Review Fall 1992, 3–12.

Charles River Associates, Inc. and Market Design, Inc., *Final Report on PPA Auction Design,* as members of the Alberta Independent Assessment Team, submitted to the Alberta Department of Resource Development, December 30, 1999.

Chuang, A. and Wu, F., *Capacity payments and pricing of reliability in competitive generation ma*rkets, Proceedings of the 33rd Hawaii International System Science Conference, January 2000.

Cramton, P. and Wilson, R., *A Review of ISO New England's Proposed Market Rules,* Market Design Incorporated, September 9, 1998.

Cramton2, P., and Lien, J., *Eliminating the Flaws in New England's Reserve Markets,* working paper, New England ISO, March 2, 2000.

DePillis, M., *A Demand Curve for Operating Reserves,* a PowerPoint presentation, New England ISO, January 8, 2000.

De Vany1, A. S., *Electricity Contenders: Coordination and Pricing On An Open Transmission Network,* Regulation, published by the CATO Institute, 1999.

De Vany2, A. S. and Walls, D., *Open Transmission and Spot Markets for Power: Models of Power and Transmission Pricing on the Western Network,* Institute for Mathematical Behavioral Sciences, working paper MBS97-07, University of California, Irvine, 1997.

Domowitz, I., *Automating the Continuous Double Auction in Practice: Automated Trade Execution Systems in Financial Markets,* in The Double Auction Market: Institutions, Theories, and Evidence, Friedman, D. and Rust, J., eds., New York: Addison-Wesley Publishing Company, 1993.

Electric Reliability Council of Texas, Inc., *ERCOT Protocols,* filed with the Texas Public Utilities Commission, January 5, 2001.

Fabra, N., *Uniform Pricing Facilitates Collusion: The Case of Electricity Markets,* working paper, European University Institute, October 1999.

Federal Energy Regulatory Commission, 18 CFR Part 35, Docket No. Rm99-2-000; Order No. 2000, Regional Transmission Organizations, Issued December 20, 1999, 238.

Feldman, R. A. and Mehra, R., *Auctions Theory and Applications,* IMF Staff Papers, September 1993, 485–511.

Feldman, R. A. and Mehra, R., *Auctions: A Sampling of Techniques*, Finance Development, September 1993, 32–35.

Friedman, D. and Rust, J., eds., *The Double Auction Market: Institutions, Theories, and Evidence*," Proceedings of the Workshop on Double Auction Markets held June, 1991, Santa Fe, New Mexico. Addison-Wesley Publishing Company, 1991.

Friedman, M., *How to Sell Government Securities,* Wall Street Journal, August 28, 1991, A8.

Griesmer, J. H., Levitan, R. E., and Shubik, M., *Toward a Study of Bidding Processes Part IV: Games with Unknown Costs,* Naval Research Logistics Quarterly, 14, 1967, 415–33.

Henriques, Diana B. *Treasury's Troubled Auctions,* New York Times, September 15, 1991, 143.

Hirst, E., *Maximizing Generator Profits Across Energy and Ancillary-Services Markets,* Electricity Journal, Vol. 13, No. 5, June 2000, 61–69.

Hirst2, E. and Kirby, B., *Retail Load Participation in Wholesale Electricity Markets,* paper prepared for Edison Electric Institute, January 2001.

Hirst3, E., *Real-Time Balancing Operations And Markets: Key To Competitive Wholesale Electricity Markets,* working paper, Oak Ridge, TN, January 30, 2001.

Joskow, P. L. and Schmalensee, R., *Markets for Power: An Analysis of Electric Utility Deregulation*, New York: MIT Press, 1983.

Joskow2, P. L., *Restructuring, Competition and Regulatory Reform in the U.S.*

Electricity Sector, Journal of Economic Perspectives, Summer, 1997.

Joskow3, P. L., *Deregulation And Regulatory Reform In The U.S. Electric Power Sector,* Brookings Papers, draft dated February 17, 2000.

Joskow4, P. L. and Kahn, E., *A Quantitative Analysis of Pricing Behavior In California's Wholesale Electricity Market During Summer 2000,* working paper MIT, November 21, 2000.

Kahn, A. E., Cramton, P.C., Porter, R H., and Tabors, R. D., *Pricing in the California Power Exchange Electricity Market: Should California Switch from Uniform Pricing to Pay-as-Bid Pricing?* A study commissioned by the California Power Exchange, January 23, 2001.

Ketchum, Jr., W. C., *Auction*! New York: Sterling Publishing Co., Inc., 1980.

Klemperer, P., *Why Every Economist Should Learn Some Auction Theory,* an invited paper prepared for the World Congress of the Econometric Society, August 2000.

McCabe, K. A., Rassenti, S. J. and Smith, V. L., *Designing Auction Institutions: Is Double-Dutch the Best?* Economic Journal, 1992.

McGuire, B., *Power Auctions and Intertemporal Production Costs: Dealing With Unit Commitment,* Power Working Paper Series, PWP-054, University of California, Berkeley, October 1997.

Mead, W. J., *Natural Resource Disposal Policy: Oral Auction Versus Sealed Bids*, Natural Resources Journal, Vol. 7, April 1987, 195–224.

Mester, L. J. *Going, Going Gone: Setting Prices with Auctions,* Federal Reserve Bank of Philadelphia Business Review, March/April 1988, 3–13.

Milgrom, P., *Auctions and Bidding: A Primer,* Journal of Economic Perspectives, Vol. 3, Summer 1989, 3–22.

Milgrom2, P., *Putting Auction Theory To Work: The Simultaneous Ascending Auction,* working paper, Stanford University, May 6, 1998.

Milgrom3, P. and Weber, R., *A Theory of Auctions and Competitive Bidding*, Econometrica, Vol. 50, November 1982, 1089–1122.

Milgrom4, P. R., *The Economics of Competitive Bidding: A Selective Survey*, Social Goals and Social Organization: Essays in Memory of Elisha Puzner, Cambridge: Cambridge University Press, 1985.

Morey, M. J., *Bidding Rules in England and Wales,* paper prepared for Edison Electric Institute, October 2000.

Myerson, R.B., *Optimal Auction Design,* Mathematics of Operations Research, Vol. 6, 1981, 58–73.

Myerson2, R. and Satterthwaite, M., *Efficient Mechanisms for Bilateral Trading*, Journal of Economic Theory, Vol. *29,* April 1983, 265–81.

Nordhaus, R., Shapiro, C., and Wolak, F., *An Analysis of the June 2000 Price Spikes in the California ISO's Energy and Ancillary Services Markets,* Market Surveillance Committee of the California Independent System Operator, September 6, 2000.

Nordhaus, R., Shapiro, C., and Wolak, F., *Analysis of 'Order Proposing Remedies for California Wholesale Electric Markets (Issued November 1, 2000)',* Market Surveillance Committee of the California Independent System Operator, December 1, 2000.

NordPool, *The Organised Markets in NordPool, the Nordic Power Exchange: The Spot Market,* October 1, 1998.

Oren, S., *Capacity Payments and Supply Adequacy in Competitive Electricity Markets,* paper presented to VII Symposium of Specialists in Electric Operational and Expansion Planning, May 2000.

Ortega-Reichert, A., *A Sequential Game with Information Flow,* Chapter 8, Models for Competitive Bidding Under Uncertainty, Stanford University PhD Thesis and Technical Report No. 8, Department of Operations Research, Stanford University, 1968, 232–254.

Rasmusen, E., *Games and Information*, Oxford: Basil Blackwell, 1990.

Reinhart, V., *An Analysis of Potential Treasury Auction Techniques*, Federal Reserve Bulletin, June 1992, 403–413.

Riley, J.G., and Samuelson, W.F., *Optimal Auctions,* American Economic Review, Vol. 71, 1981, 381–392.

Robinson, M. S., *Collusion and the Choice of Auction*, Rand Journal of Economics, Vol. 16, No. 1, Spring 1985, 141–145.

St. Goar, J., *Going Dutch*, Financial World, October 27, 1992, 86.

Schweppe, F., Caramanis, M., Tabors R., and Bohn, R., *Spot Pricing of Electricity*, New York: Kluwer Academic Publishers, 1988.

Shubik, M., *Auctions, Bidding, and Markets: An Historical Sketch,* in R. Engelbrecht- Wiggans, M. Shubik and J. Stack (eds.), Auctions, Bidding and Contracting: Uses and Theory, New York: New York University Press, 1983, 33–52.

Stoft, S., *Power System Economics*, unpublished manuscript, Berkeley, CA, 2001.

Sweeting, A., *The Wholesale Market for Electricity in England and Wales: Recent Developments and Future Reforms,* MIT, Center for Energy and Environmental Policy Research, working paper, September 5th, 2000.

Thaler, R. H., *The Winner's Curse: Paradoxes and Anomalies of Economic Life,* New York: The Free Press, 1992.

U.S. Department of the Treasury, U.S. Securities and Exchange Commission, and Board of Governors of the Federal Reserve System, *Joint Report on the Government Securities Market*, Washington, D.C.: Government Printing Office, 1992.

Varian, H. R., *Economic Mechanism Design for Computerized Agents,* paper presented at the Usenix Workshop on Electronic Commerce, New York, NY, July 11–12, 1995.

Vickrey, W., *Counterspeculation, Auctions, and Competitive Sealed Tenders*, Journal of Finance, Vol. 16, March 1961, 8–37.

Vickrey2, W., *Auction and Bidding Games,* Recent Advances in Game Theory Princeton, New Jersey: The Princeton University Conference, 1962, 15–27.

Vincent, D. R., *Dynamic Auctions,* Review of Economic Studies, Vol. 57, Issue 1, 1990, 49–61.

Vogel, T. T. Jr., *Dutch Auctions Appear to Be a Mixed Blessing for U.S.*, Wall Street Journal, January 4, 1993, C1.

Von Ungern-Sternberg, T., *Swiss Auctions*, Econometrica, Vol. 58, August 1991, 341–357.

Wilson, R., *Competitive Bidding with Asymmetric Information,* Management Science, 13, 1967, 816–820.

Wilson2, R., *Competitive Bidding with Disparate Information*, Management Science, 15, 1969, 446–448.

Wilson3, R., *Auctions of Shares,* Quarterly Journal of Economics, 93, 1979, 675–689.

Wilson4, R., *Priority Pricing of Ancillary Services,* Report to the Trust for Power Industry Restructuring, Market Design, Inc., Maryland, May 1997.

Wilson5, R. *Market Architecture,* working paper, Stanford University, June 1999.

Wilson6, R., *Activity Rules for an Iterative Double Auction,* Chapter 10, K. Chatterjee and W. Samuelson (eds.), Business Applications of Game Theory, New York: Kluwer Academic Press, 1999.

Wolak, F. A. and Patrick, R.H., *The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market,* Power Working Paper Series, PWP-047, University of California Energy Institute, February 1997.

Wolfram, C. D., *Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales,* Rand Journal, Vol. 29, No. 4, Winter 1998, 703–725.

Wolfram2, C. D., *Electricity Markets: Should the Rest of the World Adopt the UK Reforms?* Power Working Paper Series, PWP-069, University of California, Energy Institute, September 1999.

Wraight, R., *The Art Game*, New York: Simon and Schuster, 1965, 1966, 110–111.

ISO-NE, *1st Rev 69 Superseding Original FERC Electric Rate Schedule No. 6, Section 17—Market Monitoring, Reporting Market Rules & Procedures and Market Power Mitigation,* November 1, 2000.

Acknowledgements

I thank Edison Electric Institute for financial support of this project. I thank Eric Hirst, Kenneth Linder and Russell Tucker for valuable advice and support throughout the course of this project. I thank Eric Hirst, Kenneth Linder, Tim McClive, Russell Tucker, and members of the EEI Economics Committee for their helpful comments on various drafts of this report.

EEI Edison Electric Institute

701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000
www.eei.org