Design of Wholesale Electricity Markets

Hung-po Chao and Robert Wilson*

Electric Power Research Institute

Draft 990101

Note: when viewing or editing this draft document

- Turn on <Tools/Track_Changes/Highlight> so that additions and corrections can be tracked.
- Use <Insert/Comment> to insert remarks and questions.

* HChao@epri.com and RWilson@stanford.edu

Design of Wholesale Electricity Markets

Hung-po Chao and Robert Wilson

Contents

An Overview of Wholesale Electricity Markets.....4
Introduction.....7
A Short History of Restructuring.....10
Developments Abroad.....10
Developments in the U.S......11
Developments in Regulatory Policy.....13
Radical Designs.....13
Pools, Exchanges, and Bilateral Markets.....14
Contract and Market Solutions to Operational Problems.....18
Centralized Systems Based on Relational Contracts.....18
Decentralized Markets.....19
Bid Formats, Market Clearing, and Prices.....21
Scope of Markets.....21
Procedural Rules.....24
Other Designs.....26
Bids and Prices for Strips*.....27
Competing Markets.....28
Pricing and the Trading Process.....29
Equilibrium Among Competing Markets.....31
Market Power of Market Makers.....32
Complexity of the Market Structure.....32
Transmission Management.....34
Congestion Pricing of Transmission.....38
The Adjustment Market.....40
Derivation of Usage Charges.....41
Summary.....42
Hedges and Insurance.....43
Energy.....43
Transmission.....44
Ancillary Services and Losses.....45
Summary.....46
Pricing of Ancillary Services*.....46
The Process of Market Clearing and the Mode of Competition.....48
Contract Commitments and Settlements.....52
Settlement Procedures.....54
The Role of Intertemporal Operating Constraints.....55
The Role of Incentives and Gaming.....56
Mitigating Factors.....57
Summary.....58

Multiple Markets and Inter-Market Efficiency 58
Parallel Markets 58
Inter-Market Efficiency 59
Limitations of Time-Differentiated Prices 62
Feasibility 64
Ancillary Services 65
Transmission 66
Energy 67
Mitigation of Market Power 69
Investments in Transmission Capacity 72
Concluding Remarks 73
Appendix A: Glossary of Acronyms 75
Appendix B: Dictionary of Technical Terms 78
Appendix C: Effects of Market Structure on Reliability Management 85
Appendix D: Comparison with Gas and Pipeline Markets 88
What's Different About Gas and Pipelines 88
The Transmission Markets 89
The Effects of Monopoly Power 91
Implications for Restructuring Electricity Markets 92
Appendix E: Electricity Restructuring Facts & Figures 94
United States 94
California 94
New England (ISO-NE) 95
New York 97
Pennsylvania—New Jersey—Maryland (PJM) 98
Other States and some General Resources 100
Other Countries 101
Argentina 101
Australia 103
Brazil 105
Canada--Alberta 105
Canada--Ontario 107
New Zealand 108
Scandinavia (NordPool) 108
Spain 109
United Kingdom 111
Appendix F: Bibliography* 113

- An asterisk * indicates an <u>un</u>finished section.
- Appendix E needs to add: (a) Chile, (b) new Amsterdam Bourse for electricity trading, (c) implications of new European Common Market requirements for each country to deregulate/restructure its infra-structure markets, (d) expansion of NordPool to include Denmark and Finland, (e) dramatic new developments in UK, see July 1998 reports by Offer and DTI.

Design of Wholesale Electricity Markets

Hung-po Chao and Robert Wilson

An Overview of Wholesale Electricity Markets

The purpose of this chapter is to provide a functional description of the tasks accomplished by wholesale electricity markets. These markets are organized differently in various jurisdictions, and procedural rules tend to differ markedly even between systems that are decentralized to the same degree, but the basic tasks remain the same. The description here supposes a decentralized design in order to isolate the separate components. To simplify we focus mainly on the supply side, with the understanding that many tasks have analogs on the demand side; e.g., the analog of incremental generation is demand curtailment. We omit the longer-term issues of planning investments in new generation and transmission capacity.

The technology of electricity generation and transmission requires that the primary task is physical control of the system. Control has three main components.

- <u>Energy.</u> Allocation of energy generation among suppliers and among demanders.
- <u>Transmission.</u> Allocation of transmission capacity among suppliers and demanders.
- <u>Reserves.</u> Provision of reserve capacities for generation and transmission to meet contingencies and to ensure the reliability and security of the system. The provision of reserves is often called ancillary services.

Each component has an important time dimension that differentiates between forward planning and real-time operations. Forward plans distinguish between long-term and short-term, with the short-term identified by "day-ahead" plans for the next day's operating cycle and "hour-ahead" plans for the next few hours – often 1, 2, or 4 hours. Real-time operations are conducted on a time frame that ranges from 1 hour down to 5 minutes ahead of actual dispatch, and even 1 second in the case of following variations in the load.

The ultimate authority for physical control is invariably assigned to a system operator (SO) charged with responsibility for managing the transmission system. This authority includes invoking supplementary generation and reserves as needed to follow variations in the load and to maintain the stability of the transmission system. The SO's authority is virtually complete in real-time operations, but jurisdictions assign different degrees of authority over forward planning. More latitude is allowed for decentralized markets to establish plans as the time horizon increases. In highly centralized systems the SO optimizes the allocations of generation and transmission, whereas in decentralized systems the allocation is done by markets for energy and transmission that may be managed by the SO or by independent entities. In most jurisdictions the SO acquires reserves in markets that it conducts itself, although some allow participants in energy

markets to self-provide reserves. Some of the SO's operating procedures regarding scheduling, and its standards for reserves and other ingredients affecting system security and reliability, are specified by regional coordinating councils (e.g., WSCC for the western states) and the North American Electric Reliability Council (NERC).
It is important to realize that the markets for transmission and reserves are both extensions of the energy market. As will be described later, excess demand for transmission is eliminated by altering the spatial pattern of generation; similarly, reserves are provided by allocating a portion of generation capacity to stand-by status or to the job of load following on a short time scale. Transmission and reserves can also be interpreted as derived demands. A scarcity of transmission capacity is revealed by initial plans for energy generation, which then requires adjustments in the spatial distribution to alleviate congestion. The quantity of reserves is usually specified as a percentage of the load, differentiated by function and time frame; e.g.,

- A percentage (e.g., 3.5% in California) is required as "spinning" reserve, defined as generation capacity that can be fully available (subject to limitations on ramping rates for thermal generators) within ten minutes and sustained for two hours. This is interpreted as incremental generation, but some systems also require a reserve of generation that can be decremented.
- An additional percentage (again 3.5% in California) is required as "non-spinning" reserve available within thirty minutes, and a further percentage is required as "replacement" reserve that can be activated to replace spinning and non-spinning reserves as these are used up.
- Additional generation capacity is required to have "black start" capabilities so that the system can be revived after a collapse.

The category of ancillary services also includes two key functions called "regulation." One is "VAR?" that provides reactive energy at key locations to maintain the phase angle of alternating current systems. The other is automatic generation control (AGC) provided by units equipped with electronic control devices that increase or decrease generation on a second-by-second basis to follow the load.

The distinction between energy and transmission markets disappears as the time scale shrinks. In real-time operations the SO relies first on units assigned to VAR and AGC services to meet small variations in load, then markets for supplementary energy to meet larger variations on longer time scales, and when these are exhausted, draws on spinning and non-spinning reserves. A fully decentralized market for supplementary energy consists of offers to increment or decrement generation, called "incs" and "decs." The SO calls on these as needed, based first on the required location, and then accepting those offered at the lowest prices. In contrast, the SO running a centralized system issues instructions for these alternations in generation patterns.

On the time scale of day-ahead planning the distinctions among energy, transmission, and reserves is stronger. In a fully centralized system the SO optimizes these components in a single joint plan, but decentralized systems tend to address them in sequence. Typically there are markets for long-term bilateral contracts as well as day-ahead markets for trades of energy deliverable in each of the hours of the next day. The aggregate of these transactions identifies the derived demands for transmission and reserves. If the demand exceeds the available transmission capacity (as specified by the SO) on any major line or interface then the line is said to be congested. When there is congestion the SO conducts a market to allocate the scarce transmission capacity among the users competing for access, and in any case the SO conducts auction markets in which it purchases reserves sufficient to maintain its reliability standards.

As mentioned, the transmission market is an extension of the energy market in which the spatial pattern of generation is altered to alleviate congestion. Some systems do not price transmission access on a spot basis, relying instead on directives from the SO that specify which generators must increment or decrement their output (called “constrained on” or “constrained off”). Those that use “congestion pricing” of transmission establish a usage charge based on the marginal cost of alleviating congestion, either on a nodal basis specific to each location, or on a zonal basis in which only congestion between major zones is priced explicitly. In a zonal system the SO uses offered bids for increments and decrements to compute the least costly way of alleviating congestion. For example, if transmission is congested between an export zone and an import zone then the SO sets the usage charge for transmission between them as the difference between its marginal cost of incrementing generation in the import zone and its marginal revenue from decrementing generation in the export zone. (An incremental bid requires payment from the SO whereas a decremental bid requires payment from the supplier.) The SO’s revenue from usage charges is usually conveyed to the owners of the transmission assets.

The market for reserves is also an extension of the energy market. For example, those generation units assigned to AGC are among those whose bids were accepted in the energy market and therefore are scheduled to operate during the next day. Spinning and non-spinning reserves are typically provided by units whose bids accepted in the energy market did not fully exhaust their productive capacity. Some sources, such as hydro and fast-start turbines, can provide spinning reserves even if not scheduled beforehand. The auction markets for reserves are unusual in their reliance on two bid prices, one for reserved capacity and one for actual generation when activated by the SO. The simplest scheme, say for spinning reserve, accepts bids solely on the basis of the offered price for reserving capacity, and pays for actual generation at the real-time spot price, using the offered energy bid only as a reservation price indicating the least price at which the owner wants to be called to generate.

The forward markets for energy and transmission are best interpreted as financial markets. The physical aspect is important, because the planned scheduling of units is based on the transactions in these markets, but in practice the physical commitment of resources is indicative rather than binding. In fact some systems settle all transactions at the real-time price, and even in those that settle day-ahead trades at the day-ahead prices it is still true that a supplier can deviate from its day-ahead schedule by paying the hour-ahead or real-time energy price for the deviation. The purely financial aspect is especially strong in the case of bilateral contracts, since often these are structured as “contracts for differences” in which the parties insure each other against differences between the real-time price and their agreed-on price, and physical differences are settled at the real-time price.

In some jurisdictions, such as Alberta and the U.K., participation in the market conducted by the SO is mandatory. Even so, bilateral trades can be accomplished privately via contracts for differences from the SO's price. In thoroughly decentralized markets the only mandatory aspect is the requirement that the SO is advised of schedules. Energy trading takes many forms, ranging from bilateral contracting mediated by brokers and dealers, to exchanges that establish clearing prices to match supply and demand. Parallel markets compete for customers based on the comparative advantages of different contracts and price-formation procedures. Those exchanges organized as public-benefit corporations are limited by their charters to market-clearing, whereas other market makers are private profit-making enterprises with considerable latitude to
Competition among market makers:
Supplementary markets: FTRs, hedges, options, ETCs, entitlements
[what next?] Anything on governance, regulation (PUC, FERC), market power

The Introduction lays out some background and issues that motivate the subsequent discussion. The following sections consider the general architecture of wholesale markets for electricity. The first examines the choice among forms of organization, such as bilateral contracting or multilateral trading, and in the latter, the choice between a market-clearing exchange or a tight pool with centrally optimized scheduling. The second examines the transmission market in some detail, and the third examines the energy market similarly. The final two sections examine linkages among multiple markets in decentralized designs, focusing on the role of contractual commitments and the requirements for inter-market efficiency.

Introduction

To establish a point of departure: the current restructuring of electricity markets is consistent with the analysis by Joskow and Schmalensee in <u>Markets for Power</u>, 1983. They foresaw competitive markets for generation, transmission facilities operated on an open-access nondiscriminatory common-carrier basis, and retail competition among power marketers that rely on regulated utility distribution companies for delivery. Regulation of the wholesale and retail energy markets would be reduced to structural requirements and operational guidelines and monitoring, while retaining substantial regulation of the "wires" markets for transmission and distribution. These changes entail unbundling energy from T&D, thereby reversing the vertical integration of utilities.

The current issues that we address here concern mainly the organization of the wholesale markets for energy and transmission, interpreted as including ancillary services and other requirements for system reliability and security. The examination of these issues in U.S. jurisdictions can benefit from the history of restructuring in the Canadian provinces, such as Alberta and Ontario, other countries such as Britain, Australia, New Zealand, and Norway, newly implemented designs in countries such as Spain, and current developments in several states in the U.S that have already implemented new market designs.[1]

[1] Two useful surveys are those prepared by Putnam, Hayes and Bartlett for the Ontario Market Design Committee, and by London Economics for the California Trust for Power Industry Restructuring.

The peculiar features of the electricity industry that must be considered include temporal and stochastic variability of demands and supplies, accentuated by the non-storability of power, multiple technologies with varying sensitivities to capital and fuel costs and environmental and siting restrictions, and dependence on a reliable and secure transmission system. The economic problems include substantial non-convexities (immobility of generation and transmission facilities, scale economies in generation, non-linearities in transmission), and externalities (mainly in transmission). As regards generation these problems have eased sufficiently in recent decades to enable competitive energy markets, but they remain important considerations in designing these markets.

The criteria for selecting among market designs include efficiency over the long term, including incentives for investment in facilities for generation and transmission. However, our exposition focuses on short-term efficiency, since this is the immediate concrete problem, and it is required for long-term efficiency. We emphasize the implications of the general principles of market design based on ideas from economics and game theory.

To motivate the subsequent sections, we describe three parts of the overall problem of market design. The basic design choice is the architecture of the market. There are many contending options.[2] The market can be centralized or decentralized; it can be based on bilateral contracting, a centralized exchange, or a tightly controlled pool; trades can be physical or financial obligations, and they can be forward or spot contracts; the market can include financial hedges or not; the "official" market can be mandatory or optional, and encourage or discourage secondary markets. As will be evident, our opinion is that on most dimensions, the purported advantage of one extreme or the other is illusory. We favor designs that mix the two extremes to capture some of the advantages of each from parallel operations. For instance, for the three time frames of long-term, day-ahead, and real-time, there are corresponding advantages from bilateral contracting, a central exchange, and tightly controlled dispatch.

After the market architecture is established, a host of details must be specified. We do not address operational aspects here. Procedural rules must be constructed carefully to suppress gaming and promote efficiency. It is not only a matter of closing all loopholes; rather, the procedural rules must solve some basic economic problems, such as effective price discovery that enables more efficient decisions by suppliers. All this presupposes that the market will be sufficiently competitive to produce an efficient outcome, so if not, then further measures are required to diminish the market power of dominant incumbents and to promote entry by newcomers. The fact that we focus on the market architecture as the basic structural decision does not mean that it should be decided first. Parallel consideration of several designs and their implementation is useful in the early stages so that their merits can be compared in light of stakeholders' interests.

[2] A balanced and useful view of these options is presented in Issue Paper 3 of the Ontario Market Design Committee, March 1998.

Our perspective is conditioned by an emphasis on strategic behavior. This seems paradoxical, since the aim is to construct a design that suppresses gaming or renders 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 key is to select a mode of competition that is most effective in realizing the potential gains from trade. To illustrate, we describe a common fallacy. It is deceptively easy to conclude that a mandatory pool based on a centralized optimization of all generation, transmission, ancillary services, etc. – as in the U.K. or PJM – can realize the full productive potential of the system. This view does not recognize that the schedules derived from an optimization program are no better than its inputs. In fact, suppliers can and do treat the program as a device whose outputs can be manipulated by the inputs they provide in the form of purported cost functions, availabilities, etc.[3] Thus, the mode of competition consists of contending efforts to influence the "bottom-line" results from the program, such as dispatched quantities and prices for energy, transmission, and ancillary services. In terms of economic theory: reliance on an optimization affects the form and strength of traders' incentives at various points in the process, but it does not obviate the role of incentives. A central design problem is to identify the best locus of incentives and competitive forces.

In addition to our strategic perspective, we appreciate that traders have practical motivations that are not included in standard economic theory. For instance, suppliers are typically skeptical of designs that make their financial viability dependent on prices derived as shadow prices on system constraints included in the formulation of an optimization program, and centrally planned operating schedules that are several steps removed from the cost data they submit. They prefer market-clearing prices derived directly from the terms they offer, and they prefer to devise their own operating schedules to fulfill offers accepted in the market. Similarly, they are leery of intrusions by the transmission system operator (SO) into the energy markets, fearing that the SO's extraordinary powers could bias the competitive process. We see two sources of these preferences. One is informational: submitted cost data is never sufficient to describe the full range of considerations relevant to a supplier. The other pertains to governance: the SO is usually described as the ISO, emphasizing its independence and adherence to operating standards derived from principles of power engineering, but few designs address the basic problem of incentives for the SO. For example, the SO is not liable for the financial consequences to traders of strict security standards that are motivated more by avoidance of any chance of mishap than an economic tradeoff between reliability and energy costs. Current designs rely on standards of transmission management inherited from the era when it was internalized within utilities who owned and operated transmission facilities for their native loads, but as this inheritance decays it will be useful to re-examine the issues of governance and incentives for the SO.

---

[3] Expositions that address these issues include Mark Armstrong, Simon Cowan, and John Vickers, *Regulatory Reform: Economic Analysis and British Experience* (MIT Press, 1994, Chapter 9); Michael Einhorn (ed.), *From Regulation to Competition: New Frontiers in Electricity Markets* (Kluwer, 1994, Chapters 2-7); and Nils-Henrik von der Fehr and David Harbord, "Competition in Electricity Spot Markets: Economic Theory and International Experience" (ISBN 82-570-9166-9, Economics Dept., University of Oslo, Norway, January 1998).

A Short History of Restructuring

As late as 1978 when the Public Utilities Regulatory Policies Act (PURPA) was passed the central role of regulated utilities was taken for granted. Even though this legislation enabled entry of independent power producers (IPP) and encouraged co-generation and renewable energy sources by imposing obligations on utilities to procure supplies from qualified facilities (QFs), it also assumed continuation of the utilities' vertical integration and monopoly franchises. Besides the evident motivations for centralized control of transmission and distribution, generation was assumed to be effectively a natural monopoly because at the time the efficient scale of a coal-fired generator, the economical technology at the time, was on the order of 1000 MW. Virtually no attention was given to retail competition as a viable option, even though a major effect of the Act was to enable major industrial customers to bypass their local utilities by purchasing from IPPs and especially from co-generators.

Five years later, Joskow and Schmalensee's book, *Markets for Power,* envisioned competitive markets for electricity, both wholesale and retail, by separating these businesses from the "wires" businesses of transmission and distribution. They foresaw presciently that the era in which scale economics justified utility monopolies in generation was passing. In the ensuing years the minimum scale of gas-fired plants declined to approximately 200 MW and efficiency increased greatly, while natural gas prices declined and wellhead prices were deregulated. From an economic viewpoint, the justification for vertical integration disappeared. On the other hand, progress on designs of efficient markets for electricity proceeded slowly and fitfully. The motive for creation of market mechanisms was strongest in those jurisdictions with high retail prices due to utilities' heavy obligations for payment of interest and principle on debt incurred in the 1970s for large nuclear and coal-fired plants, and purchases of energy from QFs under long-term contracts at prices above subsequent market prices. In California in particular the wholesale price at the California-Oregon border was usually under $20/MWh in a period when retail prices averaged $120 and even industrial sales exceeded $65.

Developments Abroad

The first changes occurred elsewhere, initially in Chile and Argentina [?], and most importantly in 1989 when the U.K. implemented an energy market for its England-Wales system. These privatizations of government enterprises were modeled on extensions of existing power pools, similar to those already in operation in various U.S. jurisdictions such as New England (NEPool), NY, and PJM. Their central features were a day-ahead optimization of unit commitments and scheduling, subject to system constraints such as transmission (although transmission was not priced on a congestion basis), based on detailed cost data submitted daily, and a uniform price for energy transactions. The U.K. design was imitated in Alberta, but not in others initiated soon after in Norway (NordPool) and Victoria (VicPool).

Norway's system had evolved slowly over two decades. Because generation came almost entirely from hydro sources there were no deep concerns about unit commitments and scheduling, nor ancillary services, so the focus was on transmission: the important innovation was an initial day-ahead energy market followed by adjustments to alleviate interzonal congestion. Subsequent integration with Sweden extended the system to include the key transmission line between the two countries, and enabled inter-country price equalization when congestion was absent, but for some years Sweden's heavily thermal system operated independently as regards scheduling, and transmission within its zone was managed by energy trading conducted by the system operator at its own expense. The eventual design of the California system incorporated various aspects: as in Norway, self-scheduling of generation and day-ahead transmission management via congestion pricing derived from the costs of adjustments; and as in Sweden, real-time load following and intra-zonal transmission management conducted by the system operator using offered incs and decs stacked in merit order. The system in Finland developed quite differently as a market for financial hedges against spot energy prices in negotiated transactions; it can be interpreted as a forerunner of bilateral markets based on contracts for differences (CFDs).

The Victoria design, implemented in stages (VicPool I, VicPool II, etc.) at roughly 2 year intervals, was an equally important model for the U.S., as were the subsequent extensions to South Australia and New South Wales and eventually the overall design for the Australian national grid (ANG) system finally implemented in late 1998. The VicPool design also relied on day-ahead markets with self-scheduling and congestion pricing of a few key inter-regional lines [?], followed by real-time markets for deviations. Like NordPool and eventually California, a key feature was multi-settlements in the sense that transactions in forward markets were settled at the market clearing prices in those markets. It seems to be the first with separate markets for ancillary services rather than reliance on reserves scheduled by the system operator.
[[Need further description; refer also to Ontario's design process. Include descriptions of new markets in New Zealand, Spain, the Amsterdam Bourse. The re-design of the Alberta system; the proposed re-design of the U.K. system.]]

Developments in the U.S.

The prospect that wholesale electricity markets, like other infrastructure markets for transport and communications, would be deregulated was evident in the U.S., anticipated by the Energy Policy Act of 1992, formalized in FERC Orders 888 and 889 in 1994 that established standards and rules for interstate commerce, especially transmission, and contemplated in California's 1993 inquiry called the Yellow Book that outlined options for new regulatory procedures, including various schemes for light-handed regulation of generation. The hiatus ended with California's April 1994 Blue Book that announced the CPUC's intention to restructure fundamentally the California markets. Although it expressed an intention to make initial decisions with a few months, it was not until December 1996 that its basic decisions were declared after legislation, mainly in the statute AB1890, provided enabling provisions. These included a four year transition period in which a non-bypassable "competitive transition charge" (CTC) was imposed to pay off the overhanging obligations of utilities for prior investments, grandfathering of PURPA-mandated contracts with QFs, basic standards for competitive energy markets (e.g., uniform hourly transaction prices for energy), management of transmission (other than those that municipal utilities chose to retain) by a system operator, and formation of the Trusts for Power Industry Restructuring to design the new system that was then to be conveyed to new public-benefit corporations – the California Power Exchange (PX) and the California Independent System Operator (ISO). These began operations in April 1998, but only after a lengthy process of approvals by FERC of governance, market designs and tariffs, complicated by a major debate among stakeholders about the designs. Until November 1996 the prototype was the U.K. system, modeled on a tight power pool with optimized scheduling and prices derived from shadow prices, but this was shelved in favor of a highly decentralized design that was assembled in the few months before the filing date of 31 March 1997 imposed by FERC. The radical features of the California design included competing energy markets, of which the PX was to be treated comparably by the ISO to all others, a separate day-ahead market for alleviating interzonal congestion using adjustment bids, separate day-ahead markets for ancillary services based on submitted offers, and a real-time market for load-following and residual intra-zonal balancing of transmission. Unsolved problems, such as contracting for reliability-must-run (RMR) units, details of the ancillary services markets, and auctions of firm transmission rights (FTRs) meeting FERC's requirements for "price certainty," continued even past the start of operations. Contracts for software were not finalized until shortly before the filing to FERC, which finally approved the basic designs and tariffs in October 1998, shortly before the intended start date of 1 January 1998 (deferred 3 months due to software problems).

In contrast, PJM began operations on schedule on 1 January 1998 using a design derived from the former tight power pool there. Its radical feature was complete unification of energy and transmission markets in a system of day-ahead nodal pricing, applied to over 2000 nodes. Initial complaints about the volatility of the nodal prices, exacerbated by a major line failure in New Jersey during a heat wave in early April and complicated by the long-standing practice of specifying contracts that allowed suppliers to inject power at nodes of their choosing, eventually subsided as the market stabilized, helped by financial markets such as the NYMEX futures markets for delivery at major hubs.[4] Nevertheless, throughout 1998 suppliers in PJM remained constrained to bid on the basis of actual costs due to FERC's continued concern about local market power at key nodes in the transmission system.

In New England, NEPool proposed establishment of a seemingly similar system operator (ISO-NE) derived from its tight power pool, although without any congestion pricing of transmission and no demand-side bidding.[5] It had the familiar ingredients of mandatory participation (including tradable obligations for sufficient installed and operable capacity), optimized unit commitments and scheduling and provisions for reserves, all settlements at the real-time price, and obligations for all suppliers to be callable for reserves and real-time operations (unless a "short-notice-transaction" such as an export to NY was submitted and accepted). Bids, however, were purely for energy so that suppliers were expected to internalize fixed-cost components, and could not be altered

[4] PJM has considered conversion of the nodal pricing system to one closer to zonal pricing by setting nodal pricing at only a few major hubs.
[5] ISO-NE absorbed the cost of out-of-merit bids to alleviate congestion, and the optimization was to be based on forecasts of hourly loads.

between the day-ahead submission and the real-time market. This initial hybrid changed before FERC's final approval in December 1998 for operations to begin in April 1999. Subsequent additions included multi-settlements, options to redeclare the costs of inc/dec adjustments before the real-time market and plans for demand-side bidding and congestion pricing of transmission, probably on a zonal basis and especially for imports and exports to adjacent control areas. These changes brought the design closer to California's except for the unified markets for energy and transmission, retention of the ISO's optimization rather than reliance on market clearing, and continuation of all suppliers' obligations to be callable for real-time operations.

[[NY, Texas, Montana, others. The crisis in the midwest in summer of 1998. Collapse of Indigo. The peculiar Midwest ISO with each utility managing its own transmission. Resistance to restructuring in some states, such as Oregon and Washington, who apparently fear equalization of their low prices with the higher prices prevailing in California. The CA auction of FTRs; PJM's issuance of TCCs.]]

Developments in Regulatory Policy

Besides operating approvals by FERC of the designs in California, PJM, and New England (the only ones approved at this writing?), the chief development in the regulatory area in 1998 was FERC's initiation of an inquiry into whether it will require system operators (now termed regional transmission operators, or RTOs) and set uniform standards, rules, and procedures.

[[Also, it removed price caps for ancillary services in CA; approved redesigned contracts and procedures for RMR units; etc.]]

[[Need here to include the substantially revised policies in U.K. ECC mandates for deregulation of infrastructure markets in the Common Market area. Changes in NordPool and its expansion.]]

Radical Designs

Because the subsequent sections concentrate on designs that are close to current norms, we first mention radical designs that are excluded. One version stems from the view that the historical importance of system reliability may be less critical with the advent of computer controlled operations. For example, the airline industry has many similarities to the electricity industry but it is organized quite differently, and the reason may be that failures or errors in a transmission grid have enormous external effects throughout the system.[6] It might be that a decade from now the best designs are more decentralized, like the airline industry, because the reliability of the transmission system can be assured without the centralized operations inherited from vertically integrated utilities. In particular, the vulnerability of the transmission system stems presently from weak monitoring and controls on injections and withdrawals, and primitive metering devices, all of which could be eliminated by technological advances. An extreme variant

[6] The similarities include economic importance and external effects, stochastic demand, capital and fuel intensity, wastage of unused capacity (because inventories are impossible), importance for efficiency of optimal scheduling, injection (i.e., takeoff and landing) charges for use of the system, the necessity of a traffic control system for safety and reliability, the high costs of failures or errors, dependence on advanced technology, etc. This analogy is due to Severin Borenstein.

imagines that the functions of the system operator could as well be managed as a franchise, provided the firm managing operations has appropriate incentives, such as liability for costs imposed on energy traders who rely on the transmission system.

Another view is that the current system designs are residues from the era of regulation in which there were inadequate incentives for product differentiation; e.g., power service differentiated by priorities or incentives for voluntary or automatic curtailment in peak periods could reduce the reliance on supply-side controls and enable more efficient investment in base-load generation facilities.

A third view is that the only unique feature of the power industry is that an optimal pricing scheme is based on congestion charges for over-demanded transmission lines, which is complicated by the implications of Kirchhoff's Laws. Organizing the entire system around this consideration seems a high cost to pay, and some argue that it would suffice to use "postage-stamp" charges for transmission, presumably differentiated by service priority, or to rely on secondary markets for trading of firm transmission rights, or even to build a transmission system sufficient to reduce congestion to a trivial minimum. This view depends on a judgment that the gains from a thoroughly optimized system for transmission and ancillary services are small compared to the gains from vigorous competition in energy markets, and in particular, avoidance of the inefficient investments (with hindsight) in generation capacity that have plagued the electricity industry over the past quarter-century.

We assume that these radical departures from current designs are not immediately relevant, if only because they imply electricity markets that are more decentralized and privately managed than is likely soon. So we focus on those design aspects that are closer to established practice.

Pools, Exchanges, and Bilateral Markets

The structural feature of broadest significance is the organization of the market. Among the myriad of possible forms, the ones most common in commodities markets are bilateral exchanges. Those organized as "rings" or "pits" depend on oral outcry of bids and asks (usually by brokers acting for traders), whereas others use computerized bulletin boards to post offers. Those that depend on market makers to establish prices are conducted by specialists who clear orders from a book or dealers who post bid and ask prices. Market makers are usual where it is important to sustain inter-temporal continuity of prices and reduce volatility, and typically they trade for their own accounts and maintain inventories. Market makers in the energy industries often play an important role reconciling differences among short and long term contracts, and more generally, providing a variety of contract forms and auxiliary services.

Compared to the other organizational forms discussed below, the most salient distinction of bilateral markets is the continual process of trading, with prices unique to each transaction. The experimental and empirical evidence indicates that in general bilateral markets are not less competitive or efficient than exchanges or pools. Among those 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.

On the other hand, bilateral markets encounter a fundamental problem maintaining efficiency in related markets for transmission and reserves. The demands for transportation and reserves are "derived" demands. For example, for each bilateral transaction the associated demand value for transmission to fulfill the contract is the sum of the two parties' gains from trade in that transaction. When parties are matched somewhat randomly into pairs for bilateral transactions, their gains from trade are also random, and thus in the aggregate express inaccurately the actual demand value of transportation. When transportation is scarce or expensive, as in the case of power transmission, market makers face a substantial task in utilizing transmission facilities efficiently. They might accomplish this by aggregating transmission demands, or by brokering transmission services, but we know of no viable theory that assures the outcome is likely to be fully efficient, taking account of the inherent externalities. Thus, on matters of efficiency in transmission, faith in purely bilateral markets requires confidence in the ingenuity of market makers. This is not necessarily an argument against bilateral markets, however, since bilateral markets can operate alongside exchanges and the system operator's market for transmission that carry more of the responsibility at the margin for insuring efficient utilization of transmission facilities.[7] The California design includes this feature, and in Scandinavia the NordPool exchange accounts for less than 20% of the market.

Exchanges and pools offer several advantages and also bring some disadvantages compared to bilateral markets. One advantage is a central market that establishes a uniform clearing price for standardized contractual commitments, and more accurately expresses the derived demand for transmission. The uniform clearing price has some minor potential to realize the last iota of the gains from trade, but often the motives are more practical.[8] For a critical commodity like electricity there is also a perceived advantage in establishing an "official" exchange with minimal transactions costs, unhindered access for all traders, transparency to enable regulatory and public scrutiny, and countervailing power against the emergence of private market makers with sufficient market power to extract some portion of the potential rents. The disadvantages lie in the reliance on restrictive contract forms and inflexible procedural rules, and if the governance structure is inadequate, some potential to dictate restrictive procedures that are more convenient for administrators than traders. In addition, most pools and exchanges rely on private bilateral markets for auxiliary services such as financial contracts to hedge prices. Attempts to maintain pools and exchanges for contracts with

---

[7] As emphasized in the Ontario Market Design Committee's Issue Paper 3, March 1998, this depends on thorough comparability in the treatment of bilateral contracts and exchange trades as regards charges for transmission and ancillary services.

[8] For instance, in California the Power Exchange's price is used to settle grandfathered contracts, and affects payments for recovery of stranded costs. Requiring the incumbent utilities to trade through the PX also makes it easier to monitor market power. In the U.K. initially and in Alberta still, hedging contracts used to mitigate the incentives of incumbents with substantial market power are based on the exchange price.

longer terms than a day ahead have mostly failed due to lack of depth and liquidity, so typically they are confined to short-forward and spot transactions.

Here we use the term exchange for a simple market clearing system for energy. Typical examples are the exchanges in Alberta and California whose functions are confined almost entirely to establishing prices for each hour that clear the forward markets for day-ahead and hour-ahead trading. Closely related are their real-time markets conducted by the system operator, who selects among those bids offered for increments and decrements in supply and demand to manage the transmission system. Exchanges can minimize transaction costs (as evident in Alberta where transaction charges are quite small) and largely preserve traders' prerogatives to determine their own scheduling. A disadvantage of an exchange confined solely to sales and purchases of energy is its separation from the transmission market. For example, in California the day-ahead energy market in the Power Exchange (PX) clears before the transmission market opens, so traders must rely on predictions about the transmission charges they will encounter later, and transmission management relies on traders' offers of incremental and decremental adjustment bids to alleviate congestion on inter-zonal lines. In some cases the exchange might be only a "pretend" market as in Alberta in its early years, where the generation and distribution subsidiaries of the major firms were so heavily hedged via mutual contracts that the exchange price was little more than a transfer price.

We use the term pool to describe a system in which participation is mandatory and the "market" includes substantial intervention into unit commitment and scheduling. Pools are carried over from the operational procedures of vertically integrated utilities who entirely managed their own generation and transmission systems to serve their native loads, for which they had regulated monopolies, and in some cases, regional "tight" power pools with full control of scheduling. Typical examples today are in the U.K. and in the northeastern U.S. (New England, New York, and Pennsylvania-New Jersey-Maryland). Pools are distinguished from exchanges by the thorough integration of the energy, transmission, and ancillary services markets, and most significantly, by a centralized optimization of unit schedules that takes account of operational considerations – not just energy generation but also capacity availability, minimum generation requirements, ramping rates, etc. – for both day-ahead scheduling and real-time operations. At the heart of such a system is a massive computer program that decides nearly all aspects of unit commitment and scheduling, usually on both a day-ahead basis and then again in real-time operations using a rolling horizon. This program is not just an OPF for energy flows but rather includes (mixed-integer, nonlinear) optimization of unit commitments schedules subject to system and security constraints.[9] A price in such a system is not a market clearing price in the usual sense that it equates demand and supply. Rather, it is obtained as the shadow price on a system constraint in an optimization program whose inputs include detailed operating specifications and bids or purported cost data. Although these prices are used for settlements *ex post* as in an exchange, they do not represent prices offered by traders.

[9] Due to the inherent complexity of this centralized optimization, such programs rely on many ad hoc techniques, so the optimization is best interpreted as an approximation.

The advantage of a pool is the tight integration of all aspects of system operations, which might enable more productive efficiency, and it is less vulnerable to imperfect links among the prices in a sequence of energy and transmission markets. Its disadvantages lie in the consequences of complete centralization, since it requires mandated participation and compliance with specified operating schedules. Suppliers (especially entrants) are often reluctant to assign the prerogatives of scheduling and some are leery of prices obtained from a computer program rather than submitted bids; indeed, they may see the program as a black box whose outputs can be affected by the cost data they submit.[10] The obligations (and penalties for noncompliance) of a supplier are substantial; e.g., acceptance of a bid in the day-ahead market implies unit commitments that typically entail availability for real-time reserve whether scheduled for reserve or not.[11] Further obligations in New England, for example, include maintenance of operable and installed capacity sufficient to cover expected loads. These may affect those pools that allow scheduling of bilateral contracts without comparable obligations; e.g., New England allows bilateral contracts to be scheduled as inflexible units, which effectively protects them from obligations to be available for reserves. Mandatory participation is a fundamental problem. It hinders development of competing markets, either exchanges or bilateral, that might prove superior or bring innovations.[12] More importantly, the rigid standards imposed to qualify for acceptance into the pool can deter or block entry of new suppliers. These standards served an important role in the regulated era but in restructured markets they could be used to maintain the dominance of incumbents.

A point to be emphasized is that the choices among these basic organizational forms are not mutually exclusive. A system that mixes forms is feasible, such as an exchange that complements a bilateral market for forward trades, followed by real-time operations managed like a pool. One justification for a mixed system recognizes the role of timing. A pool is inherently a market for physical transactions, which is appropriate and even necessary on a short time frame such as real-time operations. Exchanges and bilateral markets are essentially forward markets for financial transactions, since physical deficiencies are inconsequential and ordinarily they are settled at the subsequent spot price. Hence, the longer time frame of forward markets increases the appeal of these organizational forms.[13] It is important to recognize too that local preferences are important: the New England pool is a direct extension of the familiar tight power pool that has had operating authority there for many years, whereas in California which has never had a tight power pool the initial design based on a pool was ultimately discarded

---

[10] The prices themselves are problematic when they include, besides energy prices, subsidy payments for capacity or availability that are more easily manipulated (as purportedly has been the case in the U.K.) and that depend on arbitrary parameters such as the assigned value of lost load and an assessed probability of lost load.

[11] In New England, after the day-ahead scheduling a supplier can redeclare its bids for incs and decs, or opt out of the real-time stack of bids in merit order by submitting a short-notice transaction, if approved by the ISO, but in general all committed units are considerable available for real-time adjustments.

[12] Sources of superiority could be lower transaction costs, longer-term contracts or contracts better tailored to traders' needs, provision of auxiliary services, or differentiated products such as curtailable service or price hedges or firm transmission rights.

[13] Issue Paper 3 of the Ontario Market Design Committee, March 1998, reflects a growing consensus that a mixed system takes best advantage of the differing features of long-term bilateral contracting, forward exchange, and real-time spot markets.

in favor of a more decentralized organization.[14] And of course those parties eager to profit as market makers are advocates of bilateral markets and reluctant to compete with an exchange whose transaction costs are likely to be low.

Contract and Market Solutions to Operational Problems

It is useful to realize that differences among organizational forms stem primarily from differences in contracts and pricing. A supplier's contract with a pool is all-encompassing, comprising a bundle of long-term benefits and obligations. With an exchange the standard contract comprises only some procedural rules and short-term commitments for energy delivery at the market prices: there are few or no obligations beyond those bid for and contracted explicitly at the clearing prices in the daily and spot markets for energy, transmission, and ancillary services. And of course a bilateral contract can be customized as the parties see fit. Bidding and pricing are equally diverse. Prices set by a pool are shadow prices on system constraints and ramping limits resulting from a comprehensive optimization based on submitted bids, which may even be multi-part bids that include fixed-cost components. An energy price in an exchange merely equates demand and supply for the designated commodity, leaving to subsequent markets the determination of transmission adjustments and usage charges and reserve commitments for ancillary services. These differences reflect different methods of addressing the operational problems of electrical systems.

Centralized Systems Based on Relational Contracts

System operators established as pools are descendants of producer cooperatives run as tight power pools. Like them, pools are organized on the principles of a long-run joint venture. In addition to the usual FERC-approved forms such as tariffs, the central document is an enabling contract to which all participants subscribe. This contract specifies participants' rights and privileges as well as obligations, sanctions, and penalties. The chief obligation is assignment of operating authority to the system operator (SO). This authority is circumscribed by procedural rules and by specifications of the methods for determining prices and quantity allocations. Nevertheless, the central fact is that the SO is defined as an agent assigned the task of implementing the contract among the participants, and its objective is to serve the collective interests of the pool members. The pool membership is necessarily broader than in a producer cooperative, including demanders as well as suppliers, and more diffuse in the absence of vertically integrated utilities, so collective interests are presumably served more by efficiency than by promoting the interests of suppliers.

The pure form of a pool as solely a private contract among its participants is used in New Zealand. There, the Market Surveillance Committee (MSC) is a judicial body created by the enabling contract to enforce its terms and rules, to arbitrate disputes, and to impose penalties and sanctions. Like a supreme court, its powers extend to interpretation of the contract, establishment of precedents, and issuance of orders; and these judicial powers apply equally to the legislative branch (the pool members) and the executive branch (the

[14] Rebellious stakeholders in California occasionally referred to the pool design as Gosplan, alluding to the central plan in the former Soviet Union, whereas those in New England apparently view their tight pool as an obvious convenience.

SO).[15] The regulatory authority of the national Competition Commission is applied to monitoring the conduct of the pool to prevent abuses of market power, rather than prescribing policies to which the pool must adhere or requiring approvals of policies proposed by the pool.

Pools in the U.S. are subject to legislation and the authority of FERC to regulate interstate commerce in energy, especially transmission. They are therefore constrained by regulatory policies and orders that take precedence over contractual arrangements among the pool's members. FERC's authority is so pervasive that most of the contractual provisions must be embodied in tariffs approved by FERC. Even so, the long-term contractual relationship among the pool members remains the dominant feature, evident most clearly in the continuing obligations that membership entails. In New England, for instance, the ISO was created by NEPool, which also proposed its procedures and rules. Suppliers must meet requirements for "integration" (essentially, adequate transmission, full dispatchability, etc.), must provide sufficient installed capacity and daily operable capacity, must bid-in their entire operable capacity, and are liable for called generation whether or not assigned to and paid for reserve status – to name just a few of the obligations.[16] For a member, the features of a pool – mandatory daily participation, adherence to the SO's day-ahead unit commitments and schedules, presumed availability for reserves – are unavoidable consequences of subscribing to the pool's basic contract.

The pool organization is inherited from cooperatives among suppliers and vertically integrated utilities with much weaker conflicts of interest among members than may be the case in restructured markets. The differing interests of suppliers and demanders in competitive markets, especially as regards transmission, will likely put greater stress on the processes of governance than was usual previously. These stresses could be exacerbated if a large fraction of trades are contracted through bilateral markets: because pools such as New England allow submission of bilateral contracts for scheduling on an as-is basis without the obligations that accompany other bids, some participants can effectively opt out by choosing to trade bilaterally, leaving others to bear the burdens. So far, however, there is little evidence from which to judge the long-term viability of organizing wholesale markets as a pool.

Decentralized Markets

A nearly opposite organizational form is used in those systems that rely on decentralized markets. The SO manages the transmission grid to ensure reliable operations, including scheduling of transactions concluded in energy markets, but it buys whatever resources it needs via market purchases. Rather than invoking members' obligations as in a pool, these systems operate on the principle that essentially every service provided to the SO is priced explicitly in a market. There are no obligations beyond those contracted explicitly in a voluntary market transaction and paid for at the market prices. FERC-approved tariffs establish the basic rules of these markets, but they do not establish any long-term

---

[15] In keeping with this judicial role, the MSC's chairman is a retired judge and proceedings resemble those in a civil case in a court of law.

[16] These obligations are so extensive that reportedly the application of a new member must be approved by upwards of twelve committees of NEPool.

contractual relationship among the participants other than the terms of their daily transactions 'with the market'.

The chief example of a highly decentralized system is California. The various energy markets operate independently of the SO, and further, the Power Exchange must compete with other private exchanges and markets for bilateral contracts. The SO conducts day-ahead auction markets for adjustment bids used to alleviate transmission congestion, and for offers to provide various ancillary services. Long-term contracts are used to acquire resources needed for localized reliability needs. Even the real-time balancing market relies on voluntary adjustment bids (incs and decs) to meet contingencies, and if these are insufficient then the SO draws upon the reserves obtained in the ancillary services markets. Unlike a pool, the SO cannot invoke additional obligations except in an emergency.

The SO's job in a decentralized system can be fraught with anxiety that there will be insufficient bids. In California the SO obtained approval from FERC to impose a default usage charge when the day-ahead market for interzonal transmission adjustment bids fails; that is, when there are insufficient adjustments offered to alleviate congestion. Market failure can also occur for ancillary services, and the SO's demands for ancillary services tend to be larger to protect against failure of the real-time market. In each case, market failure is due to "bid insufficiency" as it is called in California, meaning insufficient bids offered at any price to solve the SO's operational problems. Market failure was a problem mainly in the initial months until the markets were well-established, but in the meantime a variety of devices were employed to avoid crises. The default usage charge of $250 was so onerous that the PX responded by requiring its participants to provide adjustment bids for the entire range of each unit. For a while the SO hedged against insufficient real-time offers by contracting day-ahead with importers to provide supplies if called; and it relied on large purchases of automatic generation control (AGC) for load following in shoulder hours. The SO substituted lower-priced faster-response reserves for those with insufficient or expensive bids. And to avoid reliance on decremental bids in the real-time market to offset energy from reliability-must-run units, it called these units before the opening of the PX and essentially required the supplied energy to be matched with loads contracted in the PX or bilateral markets.

Concern about the rocky start of the California system declined after the markets stabilized. Nevertheless, the experience made clear that the advantages of a decentralized system are partially offset by the risks of bid insufficiency, and by the extra costs incurred by the SO to obtain sufficient reserves to protect the reliability of the grid. The advantages are typically seen as deriving from competition among the energy market makers, and from the preference of suppliers to manage their own unit commitments and scheduling, both of which are enabled by confining the SO's role to grid management. Not enough attention is paid, however, to the hazards of market failure and the increased difficulties of grid management when the SO cannot invoke a blanket obligation of each participant to be available for contingencies encountered in real-time. The locus of the problem is the stack of adjustment bids on which the SO can draw for real-time load following and transmission balancing: in California this stack consists only of the voluntary "supplemental energy bids" and those reserves purchased day-ahead, whereas in New England the stack includes for each unit the entire range of feasible adjustments.

Bid Formats, Market Clearing, and Prices

The economic significance of the prices obtained from an energy market derives from the format in which bids are submitted and the rules for market clearing. Several of the schemes used in practice are described in this chapter. As will be seen, none is perfect. We begin by reviewing some broad issues. We then list some schemes that have been implemented, and conclude by examining innovative proposals. We focus on the supply side, but analogs on the demand-side are similar.

Scope of Markets

A primary consideration in interpreting market prices is the scope of the market. Spatial and temporal dimensions are part, but also relevant are the definition of the commodity traded and the obligations and privileges inherent in transactions.

Spatial. In many cases, a market sets a price that equates energy supply and demand in a region, usually coincident with the system operator's control area, including imports and exports, sometimes excluding municipal utilities. Examples are the initial designs in England-Wales, California, and New England. This poses two main problems, both associated with transmission.

- If transmission congestion is priced explicitly then the ultimate effect of transmission usage charges is to establish different prices for final settlements in different locations. Some systems, such as PJM establish include transmission considerations by setting a price initially at each node or hub in the grid. Others such as the California PX establish an "uncongested" day-ahead price initially and then later impose the ISO's usage charges on inter-zonal transfers to establish prices in each zone for settlements. The significance of the uncongested price is problematic, since it bears only an approximate relationship to the zonal prices used for settlements.[17]
- PJM's system of nodal prices is vulnerable to market power in localities with isolated by transmission congestion; indeed, throughout 1998 suppliers were constrained by FERC to bid on a purely cost basis, partly to prevent such market power. A similar problem occurs in California where transmission reliability requires generation from plants in particular locations. This problem is addressed by obligating such reliability-must-run plants to supply energy under long-term contracts, and beginning in 1999, the portion of this energy bid into the PX is offered on a must-take basis. The initial implementation in New England does not use congestion pricing so it relies on paying above-market bids to alleviate congestion – and the initial design makes no provisions to mitigate local market power.

The PJM system conforms to the established engineering theory of an electrical network in which the price at each node is imputed from the shadow prices on constraints in an optimized dispatch. However, the experience in both PJM and California indicate that the application of this theory must be modified in practice to address the pervasive

---

[17] After its initial year of operations, the California PX plans to redefine the PX day-ahead clearing price to be the price in a particular zone.

problems of local market power. It remains to be seen whether the New England system can replicate the operations of the previous power pool, NEPool, in which no congestion pricing is used and no protection against market power is imposed.

Temporal. The distinction between forward and spot markets is fundamental. A related aspect is the degree to which the market is a residual after previous transactions. A typical sequence consists of continual privately-run markets for bilateral contracts with negotiated prices, a day-ahead and/or hour-ahead market that sets clearing prices, and finally a real-time market that sets spot prices every few minutes. Often each market in the sequence is mainly a residual market for trading deviations from the transactions in the prior markets. Even if a bilateral contract passes through the day-ahead market the clearing price is irrelevant if the contract is written as a contract for differences (CFD). The significance of each price in the sequence depends heavily on the commitments made then.

If the day-ahead price is used only to establish day-ahead schedules and settlements are based on the spot price, as in Alberta and the initial design in New England, then its significance is hardly more than a rough predictor of the subsequent price used for settling transactions. Systems such as California settle day-ahead transactions at the day-ahead price, so the price represents accurately the financial commitments made then, and the physical commitments such as startup and ramping.[18] Nevertheless, as a forward price, it must be interpreted in terms of opportunity costs rather than actual costs, since each supplier takes account of opportunities to bid resources not committed in the day-ahead market into subsequent markets for adjustments to alleviate transmission congestion, ancillary services, the hour-ahead market, and the real-time market.

The interpretation of the spot market price depends on the market design. In California the spot market is a near-instantaneous market for residual balancing in which offers of incs and decs are accepted by the ISO to maintain the reliability of the transmission system and to follow the load – and if transmission between zones is congested, separate spot prices are established in each. In contrast, in New England an optimized dispatch for the next 24 hours is repeated every 5 minutes, so the spot price reflects ramping and startup considerations over a longer horizon looking forward. In the initial design, the startup and ramping effects over the previous hours, which affected the day-ahead price, are ignored in settlements, since they are based on only on the spot price.

Commodity Definition. The definition of the commodity traded affects the interpretation of market prices. As in the spatial and temporal dimensions described above, two basic distinctions are the place and time of delivery. Financial instruments, such as futures contracts, always make these distinctions, but it is worth remembering that, e.g., the "uncongested" day-ahead price in the initial California PX design has no spatial definition, and the day-ahead price in the initial Alberta and New England designs have little financial significance since settlements are based on later spot prices. Several other terms and conditions are relevant:

[18] California allows portfolio bidding so the actual unit commitments are not made until after the close of the market.

- Is an allowance for transmission losses included and how is it paid for? California began operations without including its "generation meter multipliers" to account for losses. Some systems include charges for losses in the general uplift charge, others charge demanders for losses. Some use a linear formula for losses, others use a quadratic formula.
- Are transmission charges included or are they imposed as surcharges later? PJM includes transmission charges in its construction of nodal prices. The California PX imposes usage fees for inter-zonal transmission as a later surcharge to construct zonal prices, and the costs of intra-zonal congestion are absorbed by the ISO, as they are in the initial New England design. Indeed, due to portfolio bidding, the California PX's uncongested price reflects no specific information about the inter-zonal flows resulting from the subsequent unit commitments, and because the PX competes with other energy markets its own flows are not the final determinants of congestion.
- Does the price reflect inclusion of ancillary services in the initial dispatch? New England schedules units day-ahead with specified allowances for reserves, which may include startups solely for this purpose. In contrast, California defers the markets for ancillary services until after the day-ahead energy and transmission markets. Customers of the ISO can pay for ancillary services or self-provide them, whereas the costs of reliability-must-run contracts are included in the uplift charge.

These are just a few of the ways in which seemingly analogous prices actually differ among jurisdictions.

<u>Obligations and Privileges</u>. Transactions in different jurisdictions are accompanied by substantially different terms and conditions. Penalties for noncompliance differ substantially. And compliance means different things: in California a day-ahead transaction is a financial contract, since deviations are simply charged at the spot price, whereas in New England a unit that is scheduled day-ahead is obliged to be available for real-time dispatch whether or not it is scheduled and paid for reserve status.[19] The California PX obligates traders to provide adjustment bids for alleviating transmission congestion but some other exchanges do not. Counter-party risk is inherent in bilateral markets, but largely absent from exchanges in which all transactions are between the exchange and a trader. Privileges vary: Alberta allows schedule adjustments up to a few hours before real-time, California's portfolio bids enable a supplier to defer unit commitments until after the day-ahead market, New England allows "short notice transactions" that enable a supplier to opt out of the real-time market, and those scheduled day-ahead are assured of operating at or above their lower operating limits (LOLs).

These categories suffice to indicate that the interpretations of prices in different jurisdictions differ appreciably depending on how the scope of market transactions is defined along several dimensions. Now we turn to the more specific effects of the procedural rules of the market.

---

[19] If the ISO approves, a supplier can remove all or a portion of a unit from reserve availability by submitting a "short-notice transaction" such as an export or shutdown.

Procedural Rules

Some markets are little affected by procedural rules. It is possible to structure a bilateral market with rules that govern bids, offers, and transactions in an electronic bulletin-board (EBB) market for standard contracts, such as peak and off-peak strips. In fact, however, most bilateral markets rely on unstructured, private negotiations of prices and other terms. At the other extreme are tight power pools that require suppliers to submit elaborate cost and operating data and then optimize scheduling and dispatch to meet the load at the minimum cost overall. Prices have quite different meanings: the shadow prices obtained from optimized dispatch need not bear much relation to prices negotiated bilaterally. Between these extremes are the procedural rules of organized exchanges.

Power exchanges trade standardized contracts for energy at uniform prices determined by equality of aggregate supply and demand. The exchange is the counter-party to each contract and also serves as the clearinghouse to settle transactions. In a few jurisdictions such as California and England-Wales the power exchange is separate from the operator of the transmission grid, but in others the system operator runs the energy market. Each price and the traded quantities are calculated by an explicit procedure from bids submitted in a standard format. However, the bid format and the computational procedure differ markedly among jurisdictions.

Optimized Schedules. Those exchanges run by the system operator (SO) typically integrate the energy, transmission, and ancillary services markets. Based on submitted bids and operating constraints, the entire system is optimized day-ahead and/or hour-ahead to determine tentative schedules that include specific dispatch instructions regarding unit commitments and ramping, and then again in real-time on a continuing basis. The exchange is interpreted as a residual market net of bilateral contracts, whose balanced schedules are accepted as is. Few have implemented demand-side bidding and most rely on point predictions of demand, so the optimization aims to minimize the total cost of serving the predicted demand net of the portion included in bilateral contracts. Central to the optimization is the presumption that the submitted bids represent actual costs. In practice, of course, the pretense of optimization does not alter the incentives of suppliers to manipulate the outcome: in England-Wales and Alberta the evidence is clear that the submitted data from large suppliers with market power deviates significantly from their actual costs and operating constraints. PJM differed during 1998 in that major suppliers were required to bid on the basis of cost estimates carried over from the era of regulation.

There are two basic bid formats for energy. In the classic England-Wales version a supplier provides a "multi-part" bid for each unit that includes – besides capacity availability (such as lower and upper operating limits) and operating and ramping constraints – startup costs, no-load costs, and a piecewise-linear schedule of marginal operating costs. This plethora of data (reportedly 51 numbers for each unit in the England-Wales system) is used in the optimization to decide which units to commit, the hours in which they operate, and their time-profile of operating rates. Newer versions such as New England omit startup and no-load costs, taking the view that suppliers can internalize these costs in their bids, although operating limits and ramping rates are still included in the optimization.

For price determination, the important consequence of optimization is that day-ahead prices in all hours are determined simultaneously, taking account of intertemporal constraints on startup and ramping.[20] In the actual calculation, the energy price in each hour is obtained from the shadow price on the supply = demand constraint for that hour. (In the jargon of linear programming, the primal problem of cost minimization is solved to obtain the unit schedules, and then the shadow prices are obtained are the associated dual variables.) If the submitted data were actual costs, the shadow price has the interpretation that it is the marginal social cost of supplying an additional unit of demand. Similarly, when transmission constraints are included, the nodal shadow price on the supply = demand constraint net of transmission flows, would be the marginal social cost of an additional unit of demand at that location.

Typically, these prices were not offered as bids by any supplier, since they incorporate the effects of system constraints and intertemporal constraints such as ramping rates. Some information about demand elasticity is implicit to the extent that curtailable demands are treated as dispatchable supply units. In a later chapter we describe in more detail the effects in real-time markets of basing peak-period prices only on a forward-looking optimization that ignores the prior costs of startup and ramping to reach the current supply configuration.

<u>Market Clearing</u>. Other exchanges such as California and NordPool operate on the principles of self-scheduling and market clearing. A supply bid is simply an offer to supply energy at any price at or above the bid price, and similarly a demand bid is an offer to take energy at any price at or below the bid price. With this interpretation, the market clearing price (MCP) is the one that equates supply and demand. No system optimization is involved, and it is each supplier's responsibility to schedule its own plants optimally to provide the energy sold. Settlements are made at the MCP in each market, so a supplier can deviate from its day-ahead schedule by trading the difference in the hour-ahead market or the real-time market.

This simplicity hides considerable complexity, however.

- In the day-ahead market, the price for each hour of delivery the next day is determined independently of the others. No account is taken of intertemporal constraints at the system level. Startup and ramping are chosen by each supplier individually. Similarly, spatial factors such as local generation required by the system operator for grid reliability are ignored in the energy exchange; the system operator obtains assurance that it can call on the requisite resources via long-term contracts.
- Transmission constraints are ignored in the exchange's initial market clearing. Congestion is alleviated by selecting among offered adjustment bids (incs and decs) in a separate after-market run by the system operator. NordPool uses the original

[20] No implementation has attempted a similar simultaneity on the demand side. Predicted demands are specified for each hour separately with no accounting for inter-hour dependencies or correlations.

energy bids as mandatory adjustment bids but California allows separate adjustment bids. Both confine day-ahead congestion management to inter-zonal transmission, and impose the marginal cost of the incs and decs exercised as transmission-usage surcharges on energy transactions. The system operator absorbs the cost of intra-zonal congestion management conducted in real-time using supplemental incs and decs offered on short notice. In California the motive for this separation is explicit: in order to encourage competing energy markets, the system operator is precluded from overt participation in the forward energy markets and conducts only the markets for transmission, ancillary services, and real-time balancing.
- Ancillary services are also ignored initially. The system operator conducts separate markets for the various reserve services, but also each exchange has the option to self-provide reserves.

From these three aspects omitted from the market clearing it is evident that the design of such exchanges is intended to isolate the energy markets as much as possible from the others. The day-ahead markets for transmission adjustments and ancillary services are conducted separately by the system operator after the close of the energy markets. Only the residual energy market for real-time balancing and load following is conducted directly by the system operator. One motive is to encourage competing energy markets, including multiple exchanges and bilateral contracting markets, and other motives include preferences for self-scheduling and self-provision of reserves. Implicit is an aversion to energy prices derived as shadow prices from an optimization based on purported costs and that depends on intertemporal constraints on startups and ramping that suppliers think they can manage on their own.

For the system operator, the challenge of maintaining the reliability and security of the transmission grid are obviously more severe, as described in other chapters.

Other Designs

The evolution of wholesale electricity markets continues to bring innovations. For example:

- Spain's new energy exchange allows suppliers to express daily fixed-cost components such as startup and no-load costs by specifying a minimum revenue constraint. Bids whose revenue requirements are not met are automatically withdrawn until prices rise high enough to meet the revenue requirements of all remaining accepted bids.
- To address similar concerns about fixed costs, the California PX's original design (described in more detail in chapter ?) for its day-ahead market included allowance for several iterations in which, by observing the tentative clearing prices in the early iterations, suppliers with thermal units could become increasingly sure of the pattern of prices across the hours of the next day and withdraw from some or all hours if these prices did not justify startup, and suppliers with hydro units could optimize the time pattern of spills while taking account of their daily total energy constraints. This provision has not been implemented because the hour-ahead and real-time markets have provided ample flexibility.
- Recent proposals in England-Wales would completely privatize energy trading by dispensing with any central pool or exchange. Bilateral contracts would be accepted

for scheduling up to four hours ahead of delivery, after which the system operator would alleviate congestion and purchase ancillary services.

- Recent proposals in Australia would rely entirely on the real-time market to meet contingencies, and dispense with reserved capacity to provide ancillary services other than AGC.

One can expect that other new designs will bring additional innovations. We conclude by describing an innovative bid format.[21]

*Bids and Prices for Strips**

Every power exchange sets a forward-market clearing price for each time interval, usually an hour or shorter. This time-of-day pricing of energy is adapted to the typical daily cycle of energy demand.[22] The time scale is coarser than the 5 or 10 minute intervals in which spot prices are set, but the principle that prices are differentiated by time of delivery is the same. A substantial portion of the longer-term forward contracts traded in bilateral markets use a coarser time scale that divides the day between peak and off-peak. One version used in California specifies a 16 hour peak period and an 8 hour off-peak period, and for week, a peak-period contract includes 5 weekdays of 16 hours each. Such contracts are called "strips" to connote the consecutive hours of delivery. They are standardized to increase the depth and liquidity of the market.

The key feature of strips is that they recognize the total daily costs of supplying energy from thermal generators, which incur startup and no-load costs and are constrained by lower and upper operating limits and limits on ramping rates – and ramping stresses the equipment. Therefore, economical operation requires that they operate at nearly constant rates over consecutive hours. The hours of operation during the day are immaterial; what matters is the duration of the run and the constancy of the production rate. These considerations are taken into account to some extent by optimized schedules but exchanges that set independent hourly clearing prices provide no direct means for suppliers to express their daily fixed costs and ramping limitations in their bids. As mentioned above, jurisdictions such as Spain (which allows a minimum revenue requirement to be specified in the bid) and the California PX (which initially considered a design with an iterative auction) have made some explicit allowance for these considerations; further, California's reliance on portfolio bids and self-scheduling by suppliers enables decentralized optimization of each supplier's unit commitments. But the basic fact remains that the bid format, the market clearing procedures, and the temporal pricing structure all work against easy inclusion of daily fixed costs in the bids, and therefore in the prices and trades.

---

[21] See Wedad Elmagraby (sp?), "…," Energy Institute, University of California, Berkeley, 1998.

[22] Tariffs and wholesale pricing in the U.S. are traditionally based on the time of day, partly reflecting the (erroneous) view of regulatory agencies that temporally differentiated prices provide stronger incentives for curtailing peak loads. In Europe and especially France it is more common to use Wright tariffs in which customers face prices that reflect long-term supply costs. A Wright tariff charges a fixed fee for each successive kW used during the year and a variable fee that is a function of the duration (number of hours) that kW is used during the year. These charges mimic the capital and operating costs of supplying energy to meet the demand represented by an annual load-duration curve. See R. Wilson, *Nonlinear Pricing*, Electrical Power Research Institute and Oxford University Press, 1993.

In principle, a well-designed market uses bid formats that enable each supplier to express through its bids its relevant costs and operating constraints. This principle is represented explicitly in the multi-part bids used for optimized schedules, but omitted in the bid formats of exchanges such as California and NordPool that allow only bids that state an offered price for each quantity in each hour, with no means of expressing intertemporal considerations such as the length of the production run. But this limitation is not necessary, as we describe below.

Consider a market in which two bid formats are allowed. One format enables bids by time of day, while a second format enables bids for strips.

- As is done currently, a supplier can submit for each hour a supply function specifying the minimum price at which it is willing to provide each quantity of energy, and similarly for demanders. Suppliers with flexible resources (hydro, combustion turbines) may prefer this format, as will demanders with time-of-day loads to serve.
- The bid format for strips consists of minimum and maximum operating rates and, for each MW of power, a supply function specifying the minimum price as a function of the duration of the run. This format is adapted to cycled thermal generators. The demand-side analog may be preferred by industrial loads requiring consecutive hours of supply.

[[Go on to describe the market clearing process and determination of prices. Or put all this in an appendix?]]

Competing Markets

The more decentralized the design the more latitude is allowed for private market-makers to compete for customers. A salient aspect of wholesale electricity markets is that different market-makers offer different types of contracts and trading arrangements; that is, they offer differentiated products. Early designs, such as the U.K., were so centralized that product differentiation was inhibited, but later designs reduce the system operator’s monopoly in favor of decentralized energy markets in which competition among market-makers plays a significant role.

The markets in which competition can occur include all three of the main components. Reliability generally and ancillary services in particular are invariably managed by the system operator (SO), but often the participants or market-makers can self-provide regulation and/or reserves rather than relying on purchases from the auction market conducted by the SO. Similarly, the SO retains management of the transmission system, and might use adjustment bids to establish congestion prices on a day-ahead or real-time basis, but when firm transmission rights (TCCs or FTRs) are issued part of the forward market for adjustments is supplanted by the secondary markets where these are traded.[23]

Here we focus on competing markets for energy trades. These markets differ along several dimensions.

---

[23] System operators use other tactics, such as constraining generators on or off to meet transmission constraints, or making market purchases and sales as in Sweden. Even so, to the extent that FTRs are issued the role of these interventions is transferred proportionately to the secondary markets.

Pricing and the Trading Process

A fundamental distinction can be drawn between markets that establish uniform clearing prices for standard transactions, and others that allow the terms and price of each bilateral trade to be negotiated separately. Power exchanges in all jurisdictions clear the aggregates of supply offers and demand bids at a single price for each time of delivery.[24] In multi-settlement systems this price for energy applies uniformly to all transactions in that market; later deviations are charged the price in a subsequent market. Other charges might be added, such as injection or usage charges for transmission access (often paid only by suppliers), charges for energy losses from transmission (…demanders), and a share of the cost of ancillary services (…demanders), but the basic principle remains that the energy price is determined by equating supply and demand in the aggregate. Several other features result from this key requirement.

- Often the power exchange has an official status specified by legislation or regulations, is organized as a non-profit corporation to serve the public interest, and is supervised by a governing board with representation from a wide spectrum of stakeholders. This feature is not uniform; e.g., in California the Automated Power Exchange (APX) is a private company with somewhat different operating procedures.
- The exchange is the counter-party to each sale or purchase, and the exchange manages a system for settling accounts and reconciling differences between transactions and metered quantities.
- The contracts are standardized and simple. A typical contract specifies only the quantity of energy to be injected or extracted in a specified hour at a particular location or zone. The contract is solely for energy, with no bundling of other aspects such as transmission, although there is an implied liability for additional charges for transmission, losses, and ancillary services. Via self-scheduling, participants assume responsibility for unit commitments and scheduling to meet physical feasibility requirements such as limitations on ramping rates.

These features typify public markets for standard contracts for delivery on a short forward basis, usually an hour of the next or current day. They have four elementary motivations and a fifth that is fundamental.

- One is to provide equal access on non-discriminatory terms, reinforced by a governance structure that represents the interests of both participants and the public.
- A second is to minimize transaction costs, and indeed exchanges such as Alberta charge as little as $.05/MWh.
- A third is to ensure a transparent market in which abuses of market power can be prevented or detected.
- Sometimes a fourth is to establish a standard price that can be used for other purposes, such as payments for deliveries under grandfathered contracts held by QFs or for other regulatory provisions.

The most fundamental motivation is to establish a market with the greatest possible liquidity, that is, a substantial volume of bids, offers, and transactions at relatively stable prices that reflect system-wide supply and demand conditions.

---

[24] A few primitive systems exclude demand bids and rely on the predicted quantity of aggregate demand, represented as insensitive to price. Price-sensitive interruptible loads are sometimes included as offers of negative supplies.

The key to liquidity is to conduct the market at the points in time when there is the strongest demand for transactions. In electricity markets these times are usually a day ahead of dispatch, an hour or two ahead, and real-time, ordered in terms of the volume of trade. Markets for longer-forward contracts, such as a month or week ahead, are often too thin to enable a liquid market and are therefore based on bilateral negotiations. Even though some suppliers and demanders plan their operations on a longer time horizon, the key fact is that most operations can be scheduled adequately on a day-ahead basis or shorter, and this shorter time frame allows much better information on which to predict the quantities and prices that will prevail in real-time. On the other hand, the wide spatial scope of a system-wide exchange brings problems, such as needs for some generators in critical locations to operate to ensure reliability of the transmission grid; the services of such reliability-must-run (RMR) generators are usually contracted long-term outside the general market for energy.

Exchanges usually compete with other energy markets that are organized quite differently. One kind competes by providing trades of standard contracts that differ from those traded in the exchange. In the U.S. the NYMEX commodity exchange maintains a market for futures contracts based on delivery at several key points, including the Oregon and Arizona borders in California and the western hub in PJM. In California there is a private market, conducted via posted bid and ask prices, for trading contracts for base-load power that are differentiated only by a peak period and an off-peak period; and some brokers arrange swaps.

Another kind is called a bilateral market because trades are negotiated directly between buyers and sellers at mutually agreed prices for customized contracts tailored to the needs of the parties. The contracts are often for long terms, involve price guarantees (such as payment for the difference between the exchange's spot price and the contracted price), and bundle other services such as transmission and ancillary services. The managers of these markets include brokers who solicit or arrange trades, possibly as simply as maintaining a bulletin board of posted bids and offers; and dealers who take positions to facilitate trades and may be counter-parties to contracts, in which case they also provide some settlement services.

Bilateral contracts account for the major share of electricity supply and demand, except in those few jurisdictions like California where the distribution companies are required to trade through the PX for the first few years. Recognizing this, the system operator's procedures allow bilateral contracts to be submitted on a non-discriminatory basis for scheduling of transmission and ancillary services.[25] Indeed, the exchange markets for short-forward contracts are best interpreted as markets for residual trades after accounting for the bulk of the trades that have been arranged bilaterally.[26] For instance, the percentage of trades through the exchange is reportedly about 16% in Norway, 8% in

---

[25] Among the exceptions are pools like New England that allow bilateral contracts to opt out of availability as reserve for real-time dispatch, whereas other traders are required to make committed units available.
[26] In jurisdictions where the system operator optimizes dispatch, bilateral contracts are purely financial (e.g. contracts for differences), since the SO musters all resources to meet the aggregate load. In tight power pools such as NEPool, the financial profit from the SO's re-optimization is distributed among all suppliers.

Sweden, and 5% in Victoria, with the complementary percentages accounted for by longer term bilateral contracts.

Equilibrium Among Competing Markets

A persistent difficulty in designing electricity markets is to predict the outcome of competition among different types of markets. A first impression is that there is an equilibrium determined by participants' preferences for the different types of contracts and pricing arrangements. For example, a small supplier with inflexible plant or binding financial requirements may prefer to contract far ahead at fixed prices, whereas a larger supplier with several plants, greater flexibility, and financial resources sufficient to tolerate the volatility of spot prices may see greater profits from daily bidding in the power exchange. There seem in fact to be natural divisions of the market according to the times before dispatch. Because markets for long-term contracts are thin and trades are sporadic, they are best organized by brokers and dealers who facilitate bilateral negotiations and transactions. In contrast, the volume of trading in short-forward contracts is sufficient to justify an organized exchange with standard contracts and uniform prices. And real-time trading at the spot price is best managed by the system operator, who uses these resources to follow the load and maintain the stability of the transmission system. Designs like California's are based on the explicit premise that competition among the various energy markets will eventually establish a stable division of the markets according to the comparative advantages of the different styles of organization – as indeed seems already seems to have been established in older markets such as NordPool and Victoria.

This impression is misleading, however, because vigorous debates focus on the issue of whether the various kinds of energy markets compete "on a level playing field," and whether the overall design enables market makers to capture rents that an ideal design would eliminate. For example, the system operator is typically enjoined to treat all energy markets comparably, and to ensure that charges for transmission and ancillary services do not favor one kind or another. The strongest injunction is in California, where the power exchange is not managed by the CA-ISO; rather, the CA-PX is an entirely separate corporate entity with no closer formal ties to the CA-ISO than any other market maker. Private market makers naturally want to prevent ties that would strengthen the CA-PX's competitive position through tighter coordination with the CA-ISO, and equally, proponents of the power exchange see its restrictions as encumbrances that impair efficiency and enlarge the shares of the other energy markets. These restrictions are significant, since the CA-PX must clear all transactions at a uniform price for standard contracts, cannot take a position like a dealer can, cannot conduct arbitrage in others' markets but they can in its market, etc. But one can also argue that these are inherent restrictions on a pure market-clearing mechanism.

Theoretical considerations suggest that these debates in the design phase will moderate over time as the comparative advantages of different market styles define their appropriate niches within the overall system. It is important to realize, however, that this evolution is strongly conditioned by other factors. One, of course, is whether the system operator has a monopoly on both the short-forward and real-time markets, as it does in essentially all designs other than California. This trend is likely to continue since other jurisdictions contemplating restructuring view California's separation of the CA-ISO and the CA-PX as an unnecessary complication and a possible source of inefficiencies – contrary to California's view that separation enables competing energy markets that ultimately improve efficiency. Another key factor is the evolution of the transmission market. To the extent that the SO issues long-term rights for firm transmission (FTRs), or rebates of congestion charges via TCCs, bilateral markets can be organized around bundled contracts for both energy and transmission in ways unavailable to a power exchange. It seems clear that eventually the power exchange must be able to offer comparable bundling if it is to compete successfully, at least in catering to the segment that prefers insurance or hedges against transmission charges and energy adjustments. Most power exchanges are based on fairly tight coordination with the SO's transmission market, since energy trades are based on either direct usage charges for transmission, or expected surcharges imposed after an adjustment market, imposed by the SO to alleviate congestion. If the importance of this coordination diminishes, and secondary markets for FTRs accomplish most of the forward allocation of transmission capacity, it seems likely that the comparative advantage of the power exchange will diminish too.

An alternative scenario is possible, as evidenced by the steady expansion of the California PX into other markets, including a secondary market for FTRs and a bulletin-board market for monthly and weekly trades, written as contracts for differences against the PX day-ahead price, of peak and off-peak strips.

Market Power of Market Makers

The demise of the public exchanges for short-forward trading need not be regretted if the other markets are sufficiently competitive and therefore presumably efficient. One concern, however, is that monitoring the market power of large suppliers is difficult in bilateral markets because the contracts and prices are not public and the process is not transparent. A further concern is that some market maker could acquire such a dominant role as to jeopardize the public interest in a critical commodity, but the easy contestability of bilateral markets is generally viewed as sufficient assurance that this outcome is unlikely.

Complexity of the Market Structure

A significant consideration in restructured electricity markets is the complexity of the task. For the system operator the additional complexity stems from incomplete control of the generation units, as compared with dispatch operations in a vertically integrated utility. Restructured markets also use transmission more for imports and exports. For market administrators there are entirely new tasks of clearing markets, establishing prices, and settling transactions. And when market makers compete for business, as in California directly and partially in other jurisdictions, marketing and product design become important. This section, however, takes the viewpoint of participants, especially suppliers, marketers, and distribution utilities who trade energy through the markets.

The simplest structure from the viewpoint of a supplier is a centralized market conducted by a system operator, who optimizes schedules and dispatches units. This design carries over the features of a tight power pool in a regulated industry. A typical example was the New England power pool (NEPool), which fathered the subsequent system operator, ISO-New England (ISO-NE). In NEPool, suppliers submitted cost data and the system operator determined schedules to meet the predicted load, based on a linear program that optimized operations over a 24 hour rolling horizon. The schedules were comprehensive, including consideration of transmission constraints, and providing ancillary services such as regulation and reserves. ISO-NE inherited this structure, with the main exception that suppliers provide bids for energy that internalize start-up, no-load, and to some extent ramping costs and constraints. The simplicity of NEPool participation can be seen from the fact that a seller and a buyer could agree on the financial terms of a bilateral contract, knowing that in actual operations the SO would re-optimize the entire system to meet the predicted loads at least total cost for generation and ancillary services subject to all system constraints; thus, a bilateral contract was fundamentally a contract for differences between the agreed price and the system spot price. The transition to ISO-NE removed two functions from the SO: now the provisions of a bilateral contract are not re-optimized, and participants trading in the SO's residual energy markets day-ahead and real-time must internalize such costs as startup in their bids.

Fully decentralized markets like California's impose larger burdens on participants. The chief source of complexity is the number of separate markets. A "portfolio bidder," one with several generation sources, must typically prepare bids for one or more of the competing day-ahead energy markets, unit commitments and schedules based on the energy market outcome, adjustment bids (incs and decs) for the day-ahead transmission market, and bids for several of the ancillary services markets. All of this is repeated on an hour-ahead basis, and then finally there is a real-time balancing market for which supplemental energy bids (again incs and decs) might be submitted. Besides the complexity of the physical scheduling of units to fulfill sales in the energy market, a supplier must deal with different and sometimes volatile prices for settlements in each of the day-ahead, hour-ahead, and real-time markets. If the proposed iterative format for the day-ahead energy auction were implemented by the PX, the complexity would escalate further since revised bids might be submitted every twenty minutes for several hours. Coping with all this requires a professional staff of engineers and trading specialists as well as considerable software support.

A predictable consequence of this complexity is that smaller suppliers may prefer to conclude long-term contracts, in effect paying a dealer or the manager of a market for bilateral contracts for participating in the various short-term markets conducted by the ISO. This became easier after January 1999(?) when the California ISO auctioned firm transmission rights (FTRs) that include both insurance against inter-zonal usage charges and scheduling priority whenever the adjustment market fails. In particular, it enables bilateral and dealer contracts to bundle energy and transmission into a single transaction. Those markets like the Power Exchange (PX) and APX that operate as market-clearing houses are vulnerable to defections because as presently structured they depend on active daily participation of each customer in at least the day-ahead energy and transmission markets. To compete in the long run the PX may also need to offer a market for longer-term bilateral trades of contracts for energy *cum* transmission.

The complexity of the ISO's transmission markets that operate via voluntary adjustment bids will likely also change dramatically. If and when FTRs are auctioned for 100% of the inter-zonal capacity, as preferred by FERC, this market will be largely supplanted by secondary markets for trading FTRs.

Transmission Management

Except in tight power pools, there is usually some separation between the markets for energy and transmission. This is partly a functional separation that isolates the complexity of transmission management from the simplicity of energy trading. It also reflects the fact that, unlike the private-good character of energy, transmission has substantial public-good aspects, pervasive externalities, and highly nonlinear behavior described by Kirchhoff's Laws. These features of transmission make the market design highly dependent on how property rights are defined.

If there were no scarcity of transmission capacity then energy markets could be conducted like other commodity markets. The fundamental problem in transmission is that real-time balancing and security requires control by a single authority that can draw on resources offered on a spot basis, or failing that, ancillary services held in reserve. Thus, real-time operations are invariably managed by a system operator (SO).[27] The design problem is therefore focused on how far to extend the authority of the SO, and in doing so, how much to rely on market processes.

One dimension is the extent of forward balancing. NordPool and California are representative of designs in which the SO clears a forward market for transmission on a day-ahead basis (and in California, also hour-ahead). Both clear on an inter-zonal basis and rely on adjustment bids (incs and decs) to alleviate congestion, imitating the procedures used by vertically integrated utilities. For the adjustment bids NordPool uses bids carried over from the energy market, whereas in California adjustment bids are voluntary and need not bear much relation to bids in the energy market.[28] Just as there is a sequence of energy prices at which transactions in the day-ahead, hour-ahead, and real-time markets are settled, so too there is a sequence of binding usage charges for transmission that apply to these transactions. Alternative schemes defer full resolution of congestion management closer to dispatch, as in recent proposals in Alberta that would defer declarations to two hours before dispatch.

Even though it is the SO who conducts the day-ahead transmission market, one motive for this market is to minimize the interventions of the SO.[29] That is, the aim is to enable a market for adjustment bids, seen as an extension of the day-ahead energy markets, to

[27] There is a distinction between the SO as the manager of a control area and the manager of the transmission system. When some assets or entitlements are owned by parties, such as municipal utilities, for whom the SO's transmission management is optional, the SO accepts the schedules they provide and they are immune to measures to alleviate congestion and immune to usage charges.

[28] To avoid problems at startup, the PX initially mandates adjustment bids, but this is a temporary measure.

[29] Another evidence of this motive is the provision that traders in the energy markets need not rely on the ancillary services acquired by the system operator, but instead can provide these themselves.

handle most transmission management by achieving inter-zonal balance before moving into same-day operations where the SO has tighter control on all aspects. This leaves the SO with what in California is called intra-zonal balancing, although in fact on short time frames it is managing the entire transmission system, as well as generation to follow loads. If the link between the day-ahead and real-time markets is sufficiently tight then the forward prices in the day-ahead markets can be expected to approximate the real-time prices, while providing a sufficient planning horizon for suppliers to schedule their units optimally.

The California system is also motivated substantially by the desire to enable competing forward markets for energy, so they must also compete equally in a forward market for transmission. This is carried to an extreme in the provision that the SO must retain the energy balance of each scheduling coordinator (SC) conducting an energy market; e.g., each inc/dec pair selected to alleviate congestion must come from the same SC. This runs some risk of short-run economic inefficiency because it does not assure equalization of the SCs' energy prices. This risk is viewed by some stakeholders as necessary to realize the longer-term benefits of vigorous competition among the SCs' energy markets, but it has been widely criticized because it lacks a clear economic justification. The partial remedy provided in California is allowance for inter-SC trades of adjustment bids, although due to its limited role as a pure market-clearing exchange the PX cannot easily participate in these trades.

At the other extreme from the NordPool and California forward markets are the designs that provide one form or another of transmission "rights" in the form of reservations, priorities, or insurance. These designs minimize the SO's role by auctioning reservations for most transmission capacity far in advance, such as six months or a year, and rely on trading in secondary markets to achieve an efficient reallocation for each hour. Those that provide physical rights encounter two fundamental problems. One is how to define and allocate rights in advance of the actual circumstances, such as loop flow that restricts capacity, or residual transfer capability enabled by the actual pattern of injections and withdrawals that occurs. The second is how long before dispatch to require release of a reservation if it is not scheduled, and setting penalties for noncompliance: if release is too close to dispatch then hoarding by a holder of an unused reservation could impair efficiency or enable one with market power to corner the market. For instance, if release can be deferred until after the day-ahead market then forward trades in that market can be impaired by hoarding of transmission capacity. If releases are frequent and substantial then the SO winds up managing transmission on a real-time basis, which can be precarious. And there is the practical difficulty that physical rights require the SO to monitor the allocation of rights to verify that submitted schedules conform to the entitlements owned. These considerations indicate that financial rights are preferable unless stringent controls on physical rights can ensure non-discriminatory open access to transmission.

Those systems that provide insurance or hedges issue transmission congestion contracts (TCCs) that reimburse the holder for the SO's transmission usage charge, or contracts for differences (CFDs) that achieve the same effect. In principle, private markets could provide such financial instruments, and so far the California design assumes they will, but other systems such as NY and PJM propose to rely on TCCs to allocate financially-firm transmission rights. A contentious issue is whether holders of TCCs should be accorded priority in scheduling when there are insufficient adjustment bids to clear the forward market for transmission. Insufficiency is seen as a possible problem because traders who are fully insured by TCCs or CFDs might have reduced incentives to provide voluntary adjustment bids, so the SO might not be able to clear the day-ahead inter-zonal market with the adjustment bids it receives, implying that inter-zonal spillovers must be alleviated in real-time by attracting sufficient resources into the (supposedly intra-zonal) imbalance market. A TCC supplemented by scheduling priority is the same as a firm transmission right for most practical purposes.

All systems that rely on voluntary forward markets for adjustments to resolve congestion are vulnerable to insufficient participation by traders, with resulting spillovers into the real-time market that might be of much larger magnitudes than this market is intended to handle. Among the measures that can mitigate this problem is a high default usage charge when the adjustment market fails to clear – a price high enough to ensure that ample resources are submitted to the real-time market. An alternative is to require adjustment bids, but this can be fruitless unless there is some assurance that they reflect accurately the traders' opportunity costs; e.g., the practice in NordPool of re-using the bids in the energy market as the adjustment bids provides stronger assurance than California's design in which the submission of adjustment bids is entirely voluntary (although a high default price when the market fails provides a strong incentive to submit bids sufficient to enable the market to clear). On the other hand, the California design enables suppliers to account for their inter-temporal operating constraints via their adjustment bids. At the heart of the California design is a free-rider problem, in the sense that each trader or market-maker can take the view that it is others' responsibility to provide sufficient adjustments to clear the market for transmission. There will be preliminary evidence about whether this problem is severe when the California market begins operations in April 1998.

A major design feature of transmission markets is the price determination process, which is closely linked to the definition of property rights. As mentioned, those systems that allocate firm transmission rights or priorities (FTRs) in advance use an auction to establish initial prices that are then updated continually in secondary markets. Such systems require the auxiliary services of a SO to establish real-time prices that exhaust the residual transfer capacity of the transmission system, but the intent nevertheless is to enable secondary markets for FTRs to allocate most of the capacity. Similarly, those that provide TCCs or CFDs to hedge transmission charges still rely on a SO for real-time operations that include setting usage charges.

In its purest form, real-time congestion pricing of scarce transmission capacity sets a usage charge for each directional link in the system, or equivalently (using Kirchhoff's Laws) an injection charge at each node. The choice between these is often based on practical considerations: there may be many more links than nodes, thereby favoring nodal pricing, but perhaps only a few links are congested recurrently, in which case link pricing is simpler.[30] More frequently, only a few major links or nodes are priced explicitly, and for forward markets it is sufficient to establish injection charges only for nodal hubs or for large zones, or usage charges for major inter-zonal interfaces as in NordPool and California.[31] These practices have important implications for the specification of rights and hedges; e.g., secondary markets are illiquid or inactive if the FTRs or TCCs are specified in point-to-point terms rather than zone-to-zone. In principle TCCs are required for every nodal or zonal pair but in practice it suffices to consider only those nominated by traders, and then issue a subset consistent with the system capacity and security constraints. Due to loop flow, a TCC can have a negative value and require the holder to pay rather than receive a usage charge; if this is impractical then the SO must absorb the cost, whereas link prices are always nonnegative.

In a competitive market, injection or usage charges are derived from the costs of alleviating congestion, not a tariff or "postage stamp" based on embedded cost. In an optimized pool the charge represents the shadow price on capacity, but in decentralized markets it represents the difference at the margin between the cost to the SO of accepting an inc (say, of supply in an import zone) and the revenue from a dec (of supply in an export zone), or the reverse in the case of a demand inc/dec pair. For example, in a two-zone situation the usage charge for the inter-zonal interface is typically the difference in terms of $/MWh between the most expensive inc in the import zone and the least profitable dec in the export zone, among those accepted by the SO. When the configuration is more complicated the SO uses an OPF program to select the bids that are accepted, taking account of loop flow and security constraints. Congestion pricing in this fashion is based on the principle that the transmission system is an open-access public facility in which (non-discriminatory) charges are imposed <u>only</u> to alleviate congestion on over-demanded interfaces. In particular, the owners of transmission assets cannot withhold capacity nor affect prices.[32]

Judging from systems in the U.S., where most transmission assets are privately owned, the typical flow of funds can be traced as follows. The SO sends the invoice for usage charges to the traders directly in the case of a pool, or to the management of an exchange (such as a scheduling coordinator (SC) in California) which then bills the traders, perhaps on a *pro rata* basis as in the PX. The payments to the SO are then conveyed to the holders of TCCs, if any, or to the owners of transmission assets to offset their revenue requirement for capital recovery. Revenue from auctions of FTRs or TCCs are similarly passed to the asset owners. In either case, the allocation among owners depends on an approximation of their revenue shares.

These schemes provide no incentives for owners to strengthen their transmission lines, which would reduce congestion rents, so the longer-term problem of congestion remains

[30] When only a few links have positive prices it is still true that nearly all nodes have nonzero injection charges.
[31] In these systems the SO operator absorbs the cost of real-time intra-zonal balancing via the imbalance market.
[32] An exception in the U.S. is that some owners of transmission assets or grandfathered entitlements, such as municipal utilities, can opt whether to assign their capacity to the SO for transmission management. If they choose not to do so, then the SO accepts their schedules without any pricing of congestion.

unsolved. Further, if the governance structure of the SO allows incumbent suppliers to veto expansion proposals, then they can foreclose opportunities to improve the competitiveness, or more accurately the contestability, of the market; indeed, it can be that all suppliers within a control area are reluctant to strengthen inter-ties that could increase imports. We know of no design presently that addresses fully the longer-term (and, due to the complex externalities and nonlinear features of transmission networks, theoretically unsolved) problem of creating incentives for efficient strengthening or expansion of the transmission system, or that collects surcharges reserved to pay for future expansion. One partial measure is that traders who build a new link to ease congestion are entitled to receive usage charges, perhaps in the form of TCCs.

Lastly, we mention a problem with transmission markets based on congestion prices. When usage charges are derived solely from the costs of alleviating congestion, traders can opt to “self-manage” congestion by curtailing their proposed power transfers sufficiently to eliminate usage charges. This is unlikely at the level of a small individual trader unless charges are imposed at the level of injection nodes or particular links. But even with large zones, market makers conducting exchanges or bilateral contracting that account for large fractions of transmission demand can self-manage in an explicit attempt to capture the congestion rents.[33] The California design encourages self-management, and indeed there is no concern about who captures the rents provided congestion is alleviated one way or another. In contrast, it is fundamental to the justification for optimized pools that all congestion rents are captured via usage charges. This depends on a naïve view of incentives and strategic behavior unless market power is so dispersed that price-taking prevails. More likely, the opportunity to capture congestion rents encourages concerted efforts to capture them.[34]

Congestion Pricing of Transmission

Pricing scarce transmission capacity in terms of the marginal cost of alleviating congestion seems mysterious to those unfamiliar with how it is done, because few other transport industries use similar methods.[35] Here we describe this technique for allocating capacity and explain its economic justification.

The rationale for congestion pricing stems from the absence of property rights; that is, the transmission system is interpreted as a public facility, akin to a highway or sea-lane, whose use is allocated on an open-access, non-discriminatory basis according to the principles of common carriage. However, unlike a bridge or road with posted tolls or tariffs, transmission access is priced on a spot basis using the minimum charges sufficient to keep demand within the limits of the capacity available. The determination of these

[33] This is not necessarily easy to do, since there is a significant free-rider problem engendered by each exchange’s preference that others bear the greater share of the burden in curtailing their aggregate transmission demands. The game is repeated daily, however, so implicit collusion is potentially feasible.

[34] Theoretical models as well as experimental results indicate that energy traders capture some portion of congestion rents, and We recall that empirical studies of the U.K. market confirm this prediction.

[35] The system operators of gas pipeline networks in Victoria, Australia, and the U.K. use congestion pricing, but it is not used in North America.

prices is more complicated than point-to-point transportation, moreover, because Kirchhoff's Laws dictate that energy flows cannot be directed.

First a few words about capacity. Transmission lines have nominal capacities based on various standard rating methods. Although these methods take account of thermal limits and stability requirements as well as a safety margin, they do not measure capacity exactly in any particular instance. In general the risk of damage increases steadily as the load increases, and in each circumstance the usable capacity is affected by environmental factors such as temperature and wind, and uncontrollable "loop flows." Moreover, because the failure of a line anywhere in the system increases flows on other lines, overall flows must be limited so that the system can withstand failure of any one line. In practice, therefore, the system operator makes predictions and judgments that establish the usable capacity of each line. Because this usable capacity depends on the entire pattern of flows in the system, it is determined by an optimal power flow (OPF) optimization program using all submitted schedules, and by a subsequent security analysis to assure that single-line failures can be sustained without harm to the remainder of the system. It is noteworthy that a line's available capacity can differ between the two directions, and further, counter-flows in one direction offset flows in the other. The fact that transmission capacity is variable and even ill-defined, and not easily divisible into shares, provides one explanation of why property rights are not always used to allocate capacity.[36] Nevertheless, to establish a standard of comparison it is worthwhile to begin by noting how systems with property rights function.

Centralized systems, such as tight power pools, operate essentially as supplier cooperatives that own or control transmission assets collectively. The system operator manages dispatch, transmission, and reserves to make best use of the resources available. For example, a typical stated objective is to minimize the aggregate cost of serving the load – meaning the predicted load day-ahead and the actual load in real-time. Some decentralized systems rely explicitly on well-defined property rights, such as firm transmission rights on specific interfaces, that are tradable in secondary markets; examples are California and PJM. Necessarily these rights are subject to curtailment by the system operator whenever transmission must be limited further for security reasons. More elaborate implementations have been proposed. For example, Chao and Peck (1990s papers) advocate a decentralized in market in which each transaction must be accompanied by purchases of capacity shares on all lines in accordance with the distribution of the flows from the injection point to the extraction point. An alternative system, called nodal pricing, implemented in PJM and proposed in NY establishes an injection/extraction charge at each point that is derived from the imputed shadow prices of line capacities and the distribution of flows across lines – the difference between these charges at the injection and extraction points is the charge imposed for the transaction. A third class of approximate pricing systems establishes prices only for a few key congested lines and from these derives the injection charges at all nodes; or sets nodal prices at a few hubs; or as in California establishes prices only for the interfaces between major zones.

---

[36] In contrast, gas pipeline capacity is fairly well-defined and stable, and correspondingly it is allocated via property rights, as described in Appendix D.

Here we focus on those decentralized systems that rely on "adjustments" to establish prices for transmission. Typical examples are those in Norway and California that use this procedure for day-ahead allocation of the transmission capacity between large zones. They differ mainly in that in Norway adjustment bids are derived from energy bids whereas in California they are separate bids and voluntary, and in Norway the zones are redefined continually to match the pattern of congestion whereas in California the zones are revised rarely. Both methods are derived from the procedures previously used in some vertically integrated utilities. The key feature is that congestion management is implemented as an extension of the energy market. That is, congestion is eliminated by altering the pattern of energy injections and withdrawals so that inter-zonal flows remain within the capacities available. California allows its customers (called scheduling coordinators) to self-manage congestion but failing that it is managed by the system operator.

The Adjustment Market

To be specific we describe the system used in California. The first step is collection of day-ahead preferred hourly schedules from participants. Using these hourly schedules, the system operator (CA-ISO) identifies whether there is excess demand for transmission across any zonal interface. If not then the preferred schedules are accepted for dispatch, but if excess demand is present then the congestion management process is implemented. This process relies on adjustment bids that were collected from participants along with their preferred schedules. Each adjustment bid is classified according to its location, and according to whether it pertains to a supply injection or a demand extraction, and further, according to whether it offers an increment or a decrement to the preferred schedule. To simplify here we describe only supply increments and decrements, noting that demand decrements and increments play analogous roles. A supply incremental bid (inc) states a range of incremental adjustment (MW over an hour) and a price ($/MWh) for a particular generator, and similarly a supply decremental bid (dec) states a range of decremental adjustment and a price for a particular generator. Note that an inc requires a payment from the SO whereas a dec offers a payment to the SO, because each is interpreted as a deviation from the energy supply commitment contracted in the prior energy market.

Using the entire collection of offered adjustments, the system operator runs an OPF optimization program designed to find the least cost way of eliminating all excess demands for transmission. The result is an "advisory re-dispatch" that recommends modifications of the participants' preferred schedules by incrementing some generators and decrementing others. A typical pattern is that in an importing zone some generators are incremented and in an exporting zone some generators are decremented.[37] In keeping with California's principle of allowing self-management of congestion, the advisory schedules can be accepted, rejected in favor of the preferred schedule, or replaced by a new schedule. In any case, whatever the responses the system operator conducts a final optimization that fixes the set of adjustments required to eliminate congestion. These

[37] The California system actually uses only matched pairs, consisting of a dec on the exporting side and an inc on the importing side of a congested interface, both from the same participant, which is one of the energy markets called "scheduling coordinators" in California.

adjustments are still voluntary, since the participants remain free to remedy deficiencies in the hour-ahead and real-time markets, but the usage charges derived from this optimization are not – they are imposed regardless, due and payable to the SO who then conveys them to the owners of transmission assets.

Derivation of Usage Charges

The usage charges, or congestion prices, are calculated as follows. For each congested interface, in the importing zone on one side there is a most expensive MWh of incremented generation, and in the exporting zone on the other side there is a least profitable MWh of decremented generation. The difference between these identifies the marginal cost to the SO of eliminating the last MWh of excess demand across that interface. This marginal cost of alleviating congestion is the usage charge imposed by the SO for all transmission across that interface. For instance, if the marginal inc accepted in the import zone is $40/MWh and the marginal dec accepted in the export zone is $30/MWh, then the usage charge is the difference $40 – $30 = $10 /MWh.

As mentioned, the Norwegian system differs in two respects, one being that the zones are defined each day from an examination of the flows implied by the preferred schedules, and other being that the original bids in the energy market are also used as the adjustment bids. The latter feature is markedly different than in California, where occasionally the supply of adjustment bids has been insufficient to eliminate congestion. In this case the adjustment market is declared to have failed and the SO imposes a default usage charge that is sufficiently high to remind participants that adjustment bids are needed.[38] Although FERC has allowed this default charge to be as high as $250/MWh, it need not be interpreted as punitive since it reflects the magnitude of the costs that might be incurred by the SO if day-ahead inter-zonal congestion that spills over into the real-time markets on a large scale must alleviated in the smaller balancing market intended only for intra-zonal congestion management and load following.

The adjustment markets in Norway and California can be interpreted as market mechanisms for accomplishing the task done by centralized systems such as New England and PJM. The net effect of the inter-zonal usage charges in California is to establish zone-specific prices for energy *cum* transmission that differ between an import zone and an export zone by the amount of the usage charge, whereas in PJM the dispatch optimization establishes an analogous location-specific price of energy *cum* transmission that is computed as a shadow price, or opportunity cost, of injections at that location.[39] PJM is actually closer to Norway in that energy bids implicitly provide the offered adjustments, and there is no fixed set of zones that limit the role of congestion pricing.

---

[38] For a supplier in an exporting zone who failed to provide decremental adjustment bids, it must be a shock when the zonal price net of the default usage charge is negative, as has happened occasionally in exterior zones exporting into California. To minimize these events the PX requires its participants to submit adjustment bids for the entire range of feasible generation of each committed generator.

[39] Not all scheduling coordinators impose the ISO's usage charge. To establish its zonal prices, the PX substitutes its own usage charge imputed from the most expensive inc-dec pair exercised among those it submitted. The PX has considered self-management of congestion but no action has been taken.

A further peculiarity of the California system is that the power exchange (CA-PX) and the system operator (CA-ISO) are separate. This requires that the invoice from the ISO for the usage charges must be allocated by the PX among its participants. This is done by imposing surcharges on energy transactions sufficient to pay the invoice from the ISO. For example, if the initial clearing price in the energy market was $20 and the usage charge is $5 then the resulting zonal prices at which transactions are settled must entail a $5 difference between the prices in the import and export zones. There are many ways this can be done, provided the allocation rule is known beforehand by participants. To illustrate, suppose that the southern zone is designated as the reference zone, meaning that it is immune to all usage charges. Then the zonal price for settling transactions in the southern zone is invariably $20, while the zonal price in the northern zone is either $20 – $5 = $15 if it is exporting, or $20 + $5 = $25 if it is importing. Clearly, traders must know beforehand which zone is the reference zone so that they can anticipate the financial consequences of inter-zonal congestion. Note further that in the northern zone the $5 surcharge when it is exporting is paid by suppliers to the PX, whereas demanders receive from the PX this amount, so that the net revenue of the PX ($5 times the difference between supply and demand in the northern zone, which is the amount exported to the southern zone) is exactly sufficient pay the invoice from the ISO.[40]

Summary

Congestion pricing imposes a transmission usage charge on each zonal interface derived from the net cost of the marginal pair of incs and decs selected to alleviate congestion. The usage charge is directional, so in effect a counter-flow receives a credit. NordPool uses the submitted bids for energy whereas California accepts adjustment bids that may differ from the energy bids and allows self-management of congestion. The rule for allocating usage charges results in a net energy *cum* transmission price in each zone.

Systems like PJM that use nodal pricing obtain the analogous result, albeit at the finer level of nodes rather than zones, from a direct optimization of dispatch that imputes shadow prices on all transmission links, from which energy *cum* injection charges at each node are derived. Rather than the 52 usage charges in California (2 directions times 26 interfaces) and 26 zonal prices, there are over 2000 nodal prices in PJM, and they are considerably more volatile.[41] As described in the next chapter, PJM provides financial instruments called TCCs that enable suppliers to hedge against this volatility. In contrast, California provides a kind of mutual insurance by aggregating nodes into large zones (or hubs) for day-ahead congestion management, the ISO absorbs the cost of intra-zonal balancing in real-time operations, and further, the ISO issues firm transmission rights (TCCs plus scheduling priority) for a portion of its inter-zonal transmission capacity.

---

[40] California's requirement for matching pairs of incs and decs from each scheduling coordinator have the further consequence that the marginal cost of alleviating congestion can differ among scheduling coordinators. Consequently, the PC can occasionally earn a "merchandizing surplus" when its own imputed usage charge is less than the one invoiced by the ISO. This occurs when the marginal pair selected by the ISO is from some other scheduling coordinator.

[41] For this reason PJM has considered converting to a system that establishes nodal prices at only a few key hubs, as few as five.

Hedges and Insurance

Various forms of insurance against price volatility play an important role in electricity markets. In some respects this is a carryover from the era of regulation in which utilities (and IPPs operating as QFs under PURPA regulations) were essentially immune to price risks. Indeed, one interpretation of the regulatory compact was that it essentially insured firms against all but imprudent actions in exchange for service obligations and supervision of investments. As firms learn to carry normal commercial risks, and recognize that annual rather than daily profit matters, the demand for insurance provisions will likely decline. It is important to recognize that pervasive insurance can severely impair efficiency by muting incentives. This is the so-called "moral hazard problem" that firms' incentives to use resources efficiently are reduced to the extent that they are protected against adverse outcomes. Transmission markets are most susceptible to moral hazard; e.g., firms insured against the usage charges imposed to alleviate congestion have few motives to provide adjustments for managing transmission. Nevertheless, the demand for insurance, especially price hedges, is pervasive and is reflected in nearly all market designs.

We distinguish here between private markets for hedges and those provisions incorporated into the market design. The private markets include energy futures contracts traded on commercial exchanges such as NYMEX, swaps and other contracts that avoid transmission charges, and privately issued transmission congestion contracts (TCCs) that provide insurance against usage charges. Private markets do not present such severe problems of moral hazard because usually the issuers of such financial instruments include safeguards; e.g., private TCCs are often restricted to actual transmission, as opposed to scheduled transmission, and therefore discourage over-scheduling of transmission that tends to drive up the day-ahead usage charges. Another important mode of insurance is long-term bilateral contracting, often in the form of a contract for differences between the contracted and actual spot prices; in effect, bilateral contracts provide mutual insurance between the seller and buyer.

There are many insurance provisions incorporated into market designs, and often they are implicit rather than explicit. Regulated rates of return for transmission owners insure them against volatile revenues and assure users of adequate investment in public-access facilities; inter-zonal transmission usage charges provide mutual insurance among those participants in the same zone; day-ahead energy markets provide insurance against real-time prices; and provision of ancillary services provides insurance against extreme spot prices. In jurisdictions such as California the system operator absorbs the net cost of real-time load-following and intra-zonal balancing. Frequently, some form of uplift distributes shared costs among all the participants, in the form of a surcharge on the price of energy. A few examples are detailed below.

Energy

The most common forms of explicit insurance against energy prices included in market designs are legislated hedges that assure fixed prices for some portion of the generation of major incumbents. These hedges are justified by their tendency to mitigate market power, since the higher hedged price discourages withholding of output that would increase the market price. They also provide reimbursement for investments undertaken in the earlier era of regulation, and by forestalling closures or mothballing of older plants they provide a contingency reserve. Prominent examples were the U.K. and Alberta in their early years. Their negative effects can include discouraging new construction and preventing entry of new firms, if entrants have unequal access to hedged contracts. For example, in Alberta the hedged contract cover of the incumbents was so large that market prices remained low and little new capacity was added. Because incumbents obtained hedges for contracts between their generation and distribution subsidiaries, entrants had few opportunities to compete for these contracts; indeed, some described the Alberta market as a "pretend" market that functioned mainly to establish a transfer price between subsidiaries of the three major firms.

The basic demand for long-term hedges against energy prices is concentrated among entrants. A new small-scale CCGT facility is typically financed using as security a long-term contract for base-load energy that covers a substantial fraction of capacity.[42] This presents a quandary, since legislated hedges are largely reserved for dominant incumbents to mitigate market power on a daily basis, whereas long-term mitigation depends on encouraging entry by reserving such contracts for newly constructed plants.

Transmission

Insurance against congestion pricing of transmission is pervasive. Jurisdictions such as the U.K. (until recently) and New England rely entirely on uniform access and transfer charges (i.e., postage-stamp pricing) that avoid congestion charges altogether and therefore, in effect, provide full insurance against congestion charges. These jurisdictions include in uplift charges the costs of scheduling and adjusting generators to alleviate congestion.

Intermediate are systems such as Australia, California, and Scandinavia, that impose congestion charges for inter-zonal transmission but not for intra-zonal congestion. In Australia the revenues from congestion charges are transferred to a fund whose shares are sold at auction; thus, by purchasing shares of the fund a participant can obtain insurance against congestion charges that functions much like a TCC. In NordPool, Norway uses inter-zonal congestion charges for zones that are defined each day to recognize the pattern of transfers, and the system operator absorbs the cost of intra-zonal adjustments to alleviate congestion. In contrast, until recently Sweden's system operator intervened directly in the energy markets to alleviate congestion, but again with the profits or losses allocated among all participants.

In California, the system operator's cost of managing real-time intra-zonal congestion (and any inter-zonal congestion that spills over from the day-ahead to the real-time market) is allocated among all participants via its general access and transaction charges. Insurance against day-ahead inter-zonal charges is available via firm transmission rights (FTRs) that include both financial protection via TCCs and physical protection via rights

[42] The physical plant is rarely considered adequate security because it is immobile and so specialized that its value in bankruptcy is largely dependent on its contract cover. This may change as electricity markets mature and active markets for generation assets develop.

to priority scheduling whenever the adjustment market fails to eliminate congestion. Quantities of 52 annual FTRs, differentiated only by the zonal interface (26) and direction (2), are auctioned periodically, beginning with an issue of one-third of capacity in 1998.

Issuance of FTRs complies with FERC Order 888 that requires each system operator to provide means for participants to ensure "price certainty" as regards transmission. FERC policy is to encourage issuance of sufficient financial instruments to cover 100% of capacity, but this policy apparently takes no account of the moral hazard problem of over-scheduling of day-ahead inter-zonal transmission. That is, a supplier insured via FTRs could submit a day-ahead schedule that produces congestion into an importing zone, thereby excluding competitors from importing there, and enables the insured supplier to obtain higher prices for its generation in the import zone. The severity of over-scheduling was closely monitored in 1999 after the CA-ISO's initial issue of FTRs for 25% of its inter-zonal capacity on a day-ahead basis.

At the other extreme are systems such as Pennsylvania-New Jersey-Maryland (PJM) that impose injection charges at every location. This is a theoretically ideal solution for spot transactions but its practical consequences are widely criticized. Injection charges that are finely differentiated according to location are highly volatile – some would say random – and usually the entire pattern of prices stems from congestion on a few key inter-ties. In principle, the TCCs issued by PJM provide insurance but secondary markets for point-to-point TCCs are inherently too thin to sustain a viable market, so further provisions are required to reconfigure the issued TCCs to account for transactions and changing patterns of power flows. There are further practical problems: due to loop flow, injection charges can be negative and TCCs can require payments from those who hold them, although in practice these payments are usually forgiven.

Other solutions have been proposed but not implemented. The most radical is the scheme devised by Chao and Peck[43] for private management of system operations as a franchise. They propose that the franchisee sell insurance to participants that repays the cost of interruptions (e.g., constrained on or off, or curtailments) imposed to maintain system balance and security. Because the franchisee compensates suppliers and demanders for the costs imposed by its actions, it has the right incentives to select an optimal set of curtailments. Such a system requires that participants' select insurance provisions that reflect accurately their costs of curtailments; one scheme that in principle could provide appropriate incentives for participants is described by Wilson in a companion article that relies on the theory of priority pricing of insurance.

Ancillary Services and Losses

Ancillary services are invariably provided on an insured basis. Some systems allow self-provision of ancillary services, but otherwise the system operator allocates the cost of those resources it acquires through general charges, usually to demanders. The contractual obligations of suppliers in pools are considerably more complicated since

[43] *Energy Journal*, January 1997.

typically if a unit is committed day-ahead then it is obliged to be available for real-time adjustments whether or not a portion of its potential capacity is specifically designated for reserve status.

Losses are also insured to a substantial degree, since often they are calculated via a simple linear formula rather than according to nonlinear (approximately quadratic) formulas that are more accurate.

Summary

Overall, wholesale electricity markets are infused with myriad forms of implicit or explicit insurance. Apparently this reflects an insistent demand from participants to avoid the volatility of the prices for various basic components, but it is likely that this demand will moderate as they become familiar with the normal range of price variation in markets that are repeated daily; indeed, this range is minor compared to most other industries, and the daily repetition provides more predictability than other commodity markets. Insurance and price hedges are normal accompaniments of commodity markets and can contribute to overall efficiency. The persistent hazard, however, is that insurance will diminish incentives for use of scarce resources in response to prices.

Pricing of Ancillary Services*

[partial extract from critique of ISO-NE design][insert new Chao-Wilson material]
The pricing of ancillary services specified in the NEPOOL Market Rules and Procedures, Section 6, is the most unusual feature of ISO-NE's proposed pricing structure. We have serious reservations that are described below.

It will suffice here to consider a thermal unit reduced from its high operating limit (HOL) to provide TMSR. Unlike a hydro unit that can submit a positive bid for reserving capacity for TMSR, a thermal unit's bid is deemed to be zero. Its payment per hour for each MW of capacity reserved for TMSR is essentially

$$2 \times \text{Max}\{\ 0\ ,\ \text{Estimated Spot Price for Energy} - \text{Submitted Bid Price for Energy}\ \},$$

and in addition it receives the actual spot price for energy produced if and when it is called for TMSR generation in real time. This reservation payment has several peculiar features:

- The estimated spot price is based on an optimization that ignores the requirements for ancillary services, so it presumes that the cost of the marginal generator is the same with and without set-asides for ancillary services.
- The factor 2 represents the fact that nearly the same payment is calculated twice.
  1. A *de facto* bid is calculated as the lost opportunity cost, interpreted as the unit's foregone profit from running below its HOL. This profit is calculated as the difference (if positive) between the estimated spot price and the unit's bid price.
  2. This cost is then computed again, but using the bid-in cost of the "next least expensive MW dispatched" in place of the estimated spot price, and again credited to the unit as an estimated "production cost change."

The *de facto* bid (1) might be an appropriate compensation for reserving capacity for TMSR. Indeed, if no formula payment were provided in the rules, then this is approximately what the unit's owner would want to bid for reserving capacity for TMSR, since it represents the profit foregone from not selling generation in the energy market. The *de facto* bid is actually only an upper bound on this foregone profit, however, since in fact the unit may also earn additional profit if it is called for TMSR generation, for which it is compensated at the spot price.

On the other hand, the "production cost change" (2) is a misnomer, since it bears no relation to any actual change in production costs. It might be considered a subsidy for standing ready to ramp up if called, but otherwise we are at a loss to see its motivation. The net effect of this double counting is that the TMSR selling price includes (besides the actual bid if the unit is hydro) the real-time price twice, once as a component of the lost opportunity payment, and again as part of the lost opportunity clearing price.
We see no economic justification for the inclusion of both the lost opportunity cost and the so-called production cost change. If we are correct in our reading of the procedural rules, then the net effect is that for reserving capacity for TSMR a thermal unit is deemed by the formula to have bid twice an estimate of what it foregoes by not selling generation directly in the energy market.

Since we are unfamiliar with the origins of this payment formula we must rely on some guesswork at this stage. The basic problem seems to arise from the fact that a thermal unit submits only a bid for energy generation. Because the ISO then selects some units for TMSR it must provide compensation for forcing the unit to operate below its HOL or desired dispatch point. This compensation is bounded above by the difference between an appropriate estimate of the spot price and the unit's bid price – if in fact it is not subsequently called for generation. The true compensation required is this upper bound less the profit from generation when actually called. If this interpretation is correct then it appears that the bonus or subsidy provided TMSR units for standing ready is the sum of the production cost change and any profits subsequently earned when called to generate. An alternative interpretation, as in the footnote, is that the production cost change is not actually provided as compensation, but is for some reason used in dispatch in order to bias the selection in favor of those units with the least opportunity cost.

The compensation for reserving capacity for TMSR that is used by other ISOs is considerably different – and settlements are much simpler because the payments rely on a uniform price rather than unit-specific payments. For example, California pays the spot price for energy generated just as does ISO-NE, but for the capacity reservation it pays only the highest bid among those selected, and rather than a *de facto* bid the unit's actual bid is used for the selection. In particular, this enables the bidder to makes its own estimate of both the foregone profit from direct energy sales, less the expected profit in real time from called generation. It should also be mentioned that most systems do not experience a shortage of resources for TMSR since in any case those units whose capacity is not fully sold in the energy market are available, and their opportunity cost is by definition zero.

Double counting can potentially produce significant price distortions. It appears optimal to underbid a unit's marginal cost of generation near the HOL so as to increase both the opportunity cost and the production change cost components. This risks being excluded from the TSMR selection, and lowers the estimated spot price, but these risks can be more than compensated by the double counting.

[Why hydro can bid but thermal cannot is unexplained, but one can surmise that hydro is being allowed the option to opt out of the selection for TMSR dispatch (by pricing itself out of the market) because of operating or intertemporal constraints, or total-energy constraints, or …., but it is peculiar because most systems think of storage hydro as ideal for TMSR. Note too that thermal units sufficiently below HOL receive **no** opportunity cost payment, if we understand correctly.]

The Process of Market Clearing and the Mode of Competition

The mode of competition is strongly affected by structural features of the market design. In this section we provide some examples in energy markets, and briefly, in markets for ancillary services and transmission.

Underlying these specific examples is the general view that incentive effects are not eliminated by one market design or another; rather, the form in which they are expressed depends on the specific features of the market structure. The advantage of a superior design derives from the extent to which it enables traders to express accurately the economic considerations important to them. 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.

The bid format is a key factor. For example, if the market is organized to provide hourly schedules and prices, then this tends to serve the interests of demanders for whom the time of power delivery is important, and suppliers with flexibility (e.g., ponded hydro), whereas it tends to ignore the considerations of suppliers from thermal sources, who are mainly concerned with obtaining operating schedules over consecutive hours sufficient to recover the fixed costs of startup and who are unconcerned about timing *per se*. Schemes have been devised that allow demanders to bid on a time-of-day basis while suppliers bid for operating runs of various durations; prices can then be stated equivalently in terms of hourly prices for demanders and duration prices for suppliers. Similarly, for ancillary services it is usually important to distinguish between availability payments for reserving capacity and payments for delivered energy when called by the system operator. Schemes have also been devised to allow bids in terms of priorities or adjustments, such as demands that are curtailable above a specified real-time price. We bypass these more elaborate schemes here in order to focus on the basic problem of clearing an hourly market for firm energy, either forward or spot.

In energy markets there is a basic distinction between static and iterative market processes. In a static design for a pooled market each trader provides a single bid, usually in the form of a demand or supply function, with or without a separate capacity bid or a minimum revenue requirement, and perhaps in the form of a portfolio bid for multiple generation sources that is only later converted into unit schedules. The static character lies in the fact that the initial market clearing is also the final one. The theory underlying a static design is the Walrasian theory of markets, in which the market finds a price that equates stated demands and supplies. The mode of competition lies in each trader's selection of the bid function it submits – which requires substantial guesswork since others' bids are unknown when the submission is made.

If the bids are purely for hourly energy then a static design can cause problems for suppliers with fixed costs and ramping constraints because the revenue may be insufficient to cover total costs. Designs of this sort therefore provide approximate remedies: the U.K. provides capacity payments and Spain allows suppliers to specify a minimum revenue requirement. Without elaborating details here, our view is that these auxiliary provisions engender as many gaming problems as they solve, and in the case of capacity payments based on an assumed value of lost load, are inherently arbitrary.
An iterative market process works quite differently, and reflects the Marshallian theory of markets. As in an auction with repeated bidding, it is those traders whose bids are at the margin who contend to get their bids accepted, and in each round they can base their bids on the tentative results from previous rounds. For example, suppose that as usual a supplier's bid is submitted as a series of steps at successively higher prices. In this case a an "extra-marginal" supplier, one with a step above the market clearing price, realizes that by reducing its price for that step it can be more competitive in the next round – thereby ejecting an infra-marginal bidder who in the next round becomes extra-marginal and therefore must itself improve its offered price. Thus, Marshallian competition works by inducing competition among those bidders whose steps are actually near the margin, in contrast with Walrasian competition in which the price offered for each step must be based on a conjecture about the competitive situation in the event that step is at the margin.

Iterative processes require procedural "activity" rules to ensure serious bidding throughout (and thus reliable price discovery) and to ensure speedy convergence, but they have the advantage of avoiding *ad hoc* measures to assure bidders' fixed costs are covered.[44] In a day-ahead auction the key feature is that an iterative process enables "self-scheduling" in the sense that each supplier can adapt its offers in successive rounds to the observed pattern of hourly prices. With good information about the prices it can obtain in each hour, a supplier with steam plants can itself decide on which units to schedule, their start times, and their run lengths. Similarly, a supplier with storage hydro sources can better tailor its releases to take advantage of the observed prices in peak periods. In the California PX this enables pure-energy portfolio bidding: only after the energy market clears do the portfolio bidders need to report to the system operator their unit schedules that provide the energy they sold. Instead of the detailed operating data required by the U.K.'s static pool to run its centralized optimization program, California's decentralized design assigns authority to the suppliers to schedule their own units to meet the commitments contracted in the energy market.

[44] The activity rules for the California PX are adapted from the FCC's auctions of spectrum licenses, which have been notably successful and are now used worldwide. The PX rules were tested in laboratory experiments at Caltech with good results, but they will not be implemented in the PX until late 1998, so there is presently no factual evidence on their performance in practice.

These considerations are not unique to the operation of markets organized as exchanges with an hourly market clearing price that applies uniformly to all trades. Most markets for bilateral trades allow a dynamic process in which bid and ask prices are posted continually, and any posted offer can be accepted at its offered price at any time. As in an exchange using an iterative market clearing process, traders can monitor the posted prices and the prices of completed transactions to obtain good information about the prevailing pattern of prices. And because the contracts are bilateral, each party can set its own schedule to fulfill the bargain. There are also designs for bilateral markets in which all contracts are tentative until the market clears, and then the same hourly prices apply to all completed transactions.[45]

The mode of competition for transmission is also affected by structural features of the market. At one extreme are systems that assign scheduling priority to those who hold firm transmission rights or reservations (FTRs). In these systems traders compete to acquire FTRs in the initial auction or in the secondary market, leaving the system operator with only residual responsibility for real-time balancing and security of the system. At the other extreme is the California system in which the system operator accomplishes day-ahead inter-zonal balancing by exercising options offered as adjustment bids by demanders and suppliers. Congestion on inter-zonal lines is alleviated by accepting sufficient bids for incremental generation and decremental demand in import zones, and decremental generation in export zones. Thus, in this system the transmission market is an extension of the energy market to remedy congestion by altering the location of generation.[46] Intermediate designs are those in which the system operator manages transmission by setting nodal (or zonal) injection charges based on an OPF program, but traders can obtain financial insurance by acquiring TCCs or CFDs that provide hedges against the charges imposed by the system operator. In those versions in which holders of TCCs are also accorded priority in scheduling transmission, they obtain the equivalent of firm transmission rights since they are immune to the risk that transmission charges are high. In this case, traders compete for TCCs in the initial auction and in secondary markets, but only for financial insurance rather than physical rights to schedule. Of these three, the first presents some obvious problems of inefficiency and market power if FTRs can be hoarded by dominant firms, and the second might be vulnerable to insufficient adjustment bids to enable the system operator to fully alleviate congestion.[47]

Ancillary services are especially sensitive to the bid format. Using spinning reserve as the example, it is clear that suppliers must be paid for capacity availability as well as energy generation. On this basis one might surmise that suppliers should bid both

---

[45] This design has been studied experimentally in the University of Arizona laboratory, but We have not seen a practical implementation.

[46] The separation between the day-ahead energy and transmission markets in California is due to the allowance for multiple competing markets for energy, which are then reconciled in the transmission market if congestion is revealed by the schedules they submit.

[47] This is not necessarily serious on a day-ahead basis, since the main effect is spillover into real-time balancing. When the California system begins operation in the Spring of 1998 it will be clearer whether ample adjustment bids are offered.

components, but this causes problems. The initial problem is that the system operator must evaluate such two-part bids by giving some weight (interpreted as the probability that spinning units will be called to produce) to the energy bid. But as in most multi-part bidding schemes, this is fraught with gaming problems; e.g., a bidder who thinks that a call is less probable than the weight used by the SO prefers to exaggerate the capacity bid and shrink the energy bid, and the opposite if a call is more probable. Thus the merit order of energy bids reveals less about actual costs of generation than expectations about the likelihood that spinning reserves will be activated. These incentive problems are alleviated when different procedures are used for bid evaluation and settlements. In the simplest scheme bids are accepted solely on the basis of the offered capacity price, and then settlements for energy generation are based on the system real-time energy price rather than the offered energy price.[48] That is, the offered energy price is interpreted only as a reserve price below which the supplier prefers not to be called. Thus, it provides a merit order for calling generation without distorting incentives. This scheme separates the competitive process into two parts corresponding to the two parts of the bid, one for capacity availability, and another for priority in being called to generate.

The argument is occasionally made that an energy exchange might as well augment each demand bid by the required proportion of ancillary services, or at least spinning reserve – just as is typically done for transmission losses.[49] This argument recognizes that on the demand side spinning reserve is a necessary complement to planned energy deliveries. It is mistaken, however, because on the supply side energy and spin are substitutes, not complements. Moreover, technologies differ considerably in their characteristics for spinning reserve; e.g., storage hydro sources and fast-start turbines are not subject to the ramping constraints and no-load costs of steam plants, but on the other hand, thermal plants can provide spinning reserve by operating below capacity. It is better therefore to establish a separate market for spinning reserves (and curtailable loads) along with other ancillary services so that these differing characteristics can be reflected in bids.
A peculiarity of some optimized pools is payment to suppliers for capacity in addition to energy, based on so-called multi-part bids that include components for both fixed costs and incremental energy costs, with compensating charges to demanders for "uplift". These are not payments for capacity reserved for ancillary services but rather for planned generation. This holdover from the era of regulation is unique to the electricity industry, which is the only one that does not expect suppliers to cover fixed costs, such as capital and maintenance, from the market price of its output. Although a long-run equilibrium in the industry implies prices in peak periods adequate to cover the costs of capacity idle in other periods, the motive for these payments is apparently the short-run concern that market-clearing prices for energy will be determined by incremental generation costs that will be insufficient to recover the costs of capital and O&M. Such an outcome is mainly a consequence of reliance in optimized pools on shadow prices that reflect only purported incremental costs, based on a parallel optimization of unit commitments that takes

---

[48] One qualification to this statement is that bids that would not be least cost for any real-time price are screened out before ordering the capacity bids in merit order.
[49] Most systems assign to suppliers an approximate cost of losses, without attempting an exact calculation. In California, for instance, a "generation meter multiplier" is assigned to each node and updated continually to account partially for losses, and the residual is absorbed by the SO.

account of start-up costs, ramping constraints, and minimum generation levels, as well as the uncertainty of demand and the imputed value of lost load.[50] Without elaborating fully here, we are skeptical of any such payment scheme that is not tied to explicit reservation of capacity, such as for ancillary services, because we see it as an open invitation for manipulation. Designs such as those in California, Scandinavia, and Australia dispense with these payments by clearing the market for energy entirely on the basis of prices offered for delivered energy, leaving scheduling decisions to suppliers. It might indeed be that prices in California will reflect only incremental costs that are insufficient to recover the O&M costs of installed units, but if so then that signals excess capacity that in the long run should be mothballed or decommissioned.

Contract Commitments and Settlements

A significant dimension of market design is the character and timing of the commitments made by participants during the market process. The most important aspects of commitment are the prices on which settlements are based. Commitments are often presumed to be physical, but in fact they are usually financial since a breach is remedied by charging the defaulting party the spot price of purchases or sales to make up the difference.

In a pure bid-ask market with bilateral contracts concluded continually this aspect is usually hidden by the prevailing presumption that each contract is an immediate commitment and settlement is based on the price agreed in the transaction. However, there also designs for bilateral markets in which all agreements are tentative until a final market clearing price is established that then applies uniformly to all contracts. Also, many commodities markets operate on the principle that long-term contracts are physical commitments, with settlements pegged to prices in spot markets (which often represent only a small percentage of transactions). One power market, in Finland, operates as a financial market in which prevailing prices for futures contracts provide the “signals” used by traders arranging bilateral contracts.

In markets organized as pools we can distinguish at least three forms. In an optimized tight pool in which traders submit purported costs and availabilities, a trader commits to accepting both the prices and the unit schedules obtained from the optimizing algorithm, possibly with penalties for noncompliance. Exchanges with self-scheduling can operate either as coordinating devices or as genuine price-setting mechanisms for forward contracts. Those that settle day-ahead contracts on the basis of later real-time spot prices (e.g., Alberta, Victoria) serve mainly to allocate supplies to demands on the basis of tentative clearing prices that are not binding for settlements. In an exchange there is a strong presumption in favor of using the final market clearing price even if several iterations are used to reach that conclusion. In the California PX, for instance, tentative clearing prices are established in each round, but only the final round’s prices are binding.

---

[50] It is also a consequence of relying entirely on supply-side management, taking demand as fixed and inelastic. At the very least comparable payments should be provided to demanders who accept curtailable or lower-priority service. Demand-side measures can reduce the probability and imputed value of lost load, and thereby the reliance on peaking capacity that is idle much of the year.

On standard economic grounds one might conclude that the only relevant price for allocative efficiency is the real-time spot price, and on that basis surmise that settlements should be based on this price – implying that earlier forward contracts are not binding as regards the nominal transaction price. However, this view ignores the substantial incentive effects. To motivate the subsequent discussion, we contrast the Alberta and California designs.

The design of the California PX may seem awkward at first, and indeed it is awkward in terms of the software required for settlements, since each MWh of energy might be assigned any one of several prices. In the PX's energy market, one clearing price is financially binding for trades completed in the day-ahead forward market, another clearing price is binding in the hour-ahead forward market, and the spot price in the real-time applies to ancillary services and supplemental energy purchased by the SO. On the other hand, the advantage of this design is that traders have an incentive to bid seriously in each of the forward markets, since the trades concluded there are financially binding at the clearing price in that market.

Alberta uses the opposite design in which all settlements are made at the final spot 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 all suppliers' preference to delay commitments until close to the time at which prices for settlement are established, so that uncertainty is reduced, and each supplier's advantage from committing last so that it can take maximal advantage of the likely pattern of prices thereby revealed. The Alberta design has also invited a kind of 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 in order to obtain the best terms. Of course the other traders in Alberta now want the same privilege.

Our opinion is that the difficulties implementing the Alberta design are intrinsic to any design in which transactions are not financially binding at the clearing price in the market in which they are made. One can argue that a sequence of binding forward prices might sacrifice some efficiency compared to one in which settlements are based on spot prices, but our view is that this sacrifice is necessary to ensure that bids are serious in the forward markets. If viable forward markets are unnecessary, as perhaps in a purely hydro system, then spot-price settlements are sufficient, but it seems to me that justifications for forward markets also justify binding transactions at the clearing prices in these markets. One must, of course, ensure that the sequence of forward markets is sufficiently contestable to enable arbitrage that keeps forward prices in line (in expectation) with subsequent spot prices.

One should also keep in mind the range of alternatives for the form of the commitment. An important distinction is between physical and financial commitment. Bilateral markets are more dependent on physical commitments if there is not a viable spot market in which to remedy deficiencies – or at least a dealer or broker who provides the remedy. Optimized pools depend to some extent on a presumed physical commitment to the dispatch schedule, since otherwise the optimization would be a useless exercise. In other pools, however, we have yet to see a cogent argument for physical commitments, as compared to financial commitments, in forward markets. Provided those who default on prior commitments are liable for making up the difference with purchases at the spot price, the incentives for compliance are sufficient. Further, due to the considerable stochastic variation in supply and demand conditions in power markets, the flexibility allowed by purely financial commitments is superior.[51]

The second distinction concerns the counter-party to a contract. Bilateral trades are contracts between the transacting parties, or perhaps with a dealer, whereas in an exchange or pool the counter-party to every transaction is the exchange; that is, suppliers sell to the exchange and demanders buy from it. Typically, the exchange defines standard contractual terms, and it administers the apparatus of settlements. There is no harm in this *per se*, but it encourages the growth of alternative market-makers who offer a greater variety of contractual terms and auxiliary services more closely tailored to the needs of select customers. A mandatory pool is naturally beset by pressures to remedy one or another perceived deficiency or favoritism in the rules and contracts, since invariably the pool's standard terms are inadequate to serve equally the diverse interests of a heterogeneous group of traders.

Settlement Procedures

A basic difference among market designs is the number of settlements for energy transactions. It might be thought that the procedure for settling accounts involves only elementary considerations of metering, accounting, billing, and payments. In fact, the procedures for financial settlement of transactions involve fundamental considerations of incentives and strategic behavior. This section elaborates the main differences between designs that use single and multiple settlements for energy. We ignore here the settlements for ancillary services, which include capacity or "opportunity cost" payments, and transmission, which includes usage charges or nodal injection charges.

The simplest design from an accounting and software perspective uses a single settlement. In Alberta, Australia, New England and other jurisdictions, this procedure relies on so-called *ex post* pricing of energy. In the initial ISO-New England (ISO-NE) design, for example, each transaction for energy is settled at a system price that depends on the delivery time, not the transaction time. Thus, both day-ahead and spot transactions are settled at the spot price. This price is the system marginal cost ($/MWh) of supplying an incremental MWh calculated at five minute intervals. The calculation relies on linear programming optimization software to derive an optimal plan of dispatch over the next

[51] The California PX allows portfolio bids for energy, which do not require specific unit commitments. This provision provides more flexibility to suppliers with many plants, so it might be construed as favoring larger firms, but it is also true that smaller single-plant suppliers can band together to submit portfolio bids.

24 hours, and the price is obtained as the shadow price on the current 5-minute energy balance constraint. The optimization also depends on demand projections and suppliers' submitted bids, system limits represented by constraints on transmission and ramping, and the current operating rates of the running generators. The accounting simplicity stems from the fact that a single price applies to each metered MWh.

At the other extreme are multi-settlement designs. In California the CA-PX settles day-ahead transactions at the prices that clear the forward markets for each hour of the next day, then it settles hour-ahead transactions at the price clearing that market; and finally, the CA-ISO settles spot transactions at the spot price that clears the real-time balancing market for supplemental energy. This design requires more complicated software since several prices can apply to each metered MWh. There are two motivations for multiple settlements, one concerned with intertemporal operating constraints, the other with incentives and gaming, that are examined below.

The Role of Intertemporal Operating Constraints

Generation is subject to several intertemporal constraints on operating schedules, such as total energy for stored hydro sources, flow links among reservoirs in cascades, and for thermal sources, minimal requirements for startup and cooling, and maximum ramping rates.[52] We use ramping rates to illustrate.

A single-settlement design like ISO-NE depends on an operational protocol that initially allocates generation resources on a day-ahead basis to meet predicted demands, and then re-allocates on a spot basis to meet actual demand and supply conditions. (ISO-NE allows suppliers to redeclare their inc/dec bids after the day-ahead scheduling, so supply conditions can also include cost elements.) The software that does the initial optimization views all generators as flexible within the limits of their reported ramping rates. On the other hand, the spot re-allocation takes the operating levels of the running generators as fixed data, and adjusts them as needed within the limits of their ramping rates, based on the merit order of their inc/dec bids. This protocol has three effects on prices.

- If there were no errors in predicting subsequent demand and supply conditions then the initial and spot optimizations will produce the same shadow price for energy in each interval, so it is immaterial whether one or two settlements are used. Inaccurate predictions can, however, produce differences between the prices computed initially and later. Designs with multiple settlements recognize this difference, and often suppliers prefer an initial settlement so that they commit their units day-ahead on the basis of the prices prevailing then.
- A second effect is that peak-period prices tend to be depressed when a single settlement is used. This effect occurs because expensive generators are ramped up in earlier periods in anticipation of the peak, thereby raising pre-peak prices, but when the peak arrives the spot re-allocation takes as given the operating levels of these generators and therefore does not recognize the costs incurred earlier to achieve these levels. These effects tend to cancel out over the daily cycle. Nevertheless, suppliers

[52] We know of no system that takes account of the O&M costs of ramping, such as those due to stresses on turbines and boilers.

with more flexible units can object on the grounds that at the peak they are not receiving the price that would reflect the total cost of meeting demand. Designs with an initial day-ahead settlement make payments on the basis of suppliers' overall plans for the next day's schedule, in which they internalize the full costs of ramping before and after peak loads. Deviations from these plans are then priced separately in the real-time market.

- The third effect is that suppliers for whom ramping is costly have an incentive to re-declare their inc/dec bids to put them out of range in the merit order for re-dispatch at altered generation levels. Correspondingly, those with flexible units that can ramp quickly and cheaply can obtain premium prices in the real-time market because the inflexible units have opted out.

The limitations of the ISO-NE protocol can be interpreted as consequences of an incomplete optimization. Instead of relying on continual re-optimization based on point-estimates of predicted demand over the rolling horizon of 24 hours, the software could use the technique of linear programming under uncertainty (Dantzig, 1976?). This technique models system operations for an entire daily cycle in terms that include both initial commitments and subsequent adjustments contingent on each possible realization of demand, subject to ramping constraints that link the dispatch periods of the day. In this formulation, multiple settlements are intrinsic, since one set of prices apply to initial commitments and another set to adjustments in each contingency.

This is basically the design used in California, although it relies on repeated market clearings rather than explicit optimization. The clearing prices used for settling day-ahead transactions reflect commitments made then, whereas different clearing prices are used for the hour-ahead and real-time transactions that adjust these commitments on the basis of new information.

The Role of Incentives and Gaming

Multiple-settlement designs have the obvious advantage that suppliers' bids in each market are presumably serious, since they represent genuine commitments. These commitments in the forward markets are basically financial rather than physical, because there remain opportunities to adjust physical schedules via auxiliary trades in later markets, either hour-ahead or spot.

Single-settlement designs encounter several problems. One is that spot transactions, which typically represent less than 3%, nevertheless determine the settlement prices for all earlier transactions. Thus, volatility in the spot market can produce large swings in the total revenues obtained by suppliers. A second is the incentive to drive up the spot price, even if withholding supply is sub-optimal in that market, to increase revenues from earlier transactions. Systems such as Alberta, the U.K., and ISO-NE suppress this tactic by requiring day-ahead schedules to be binding in real-time operations, but this restriction reduces the efficiency of the response to altered supply conditions. There are also insistent pressures to circumvent the restrictions. For instance, Alberta allows importers and exporters to submit multiple schedules, of which all but one can be withdrawn depending on the contingency that is realized; further, pressure from other suppliers has led to permission to re-declare their schedules a few hours before dispatch. Similarly, ISO-NE allows imports and exports on short notice. Because a large export transaction can drive up the spot price that is used for settling day-ahead commitments, this provision creates an incentive to export solely for the purpose of improving revenues from earlier transactions. The third and most fundamental problem is that the seriousness of the day-ahead bids is jeopardized. In the initial New England design, for instance, the day-ahead bids are relevant only to determine day-ahead unit commitments and schedules but have no direct impact on financial settlements. The weakness of the link between physical schedules and financial settlements is compounded by each supplier's option to redeclare its bids for incs and decs from its initial schedule – in extreme cases this could produce spot prices that bear little relation to the supposed day-ahead prices on which the initial schedules were based.

Mitigating Factors

The preceding discussion indicates that designs with multiple settlements have some superior properties, but it is useful to realize that in practice there are factors that modify this conclusion.

One important consideration is that private markets for bilateral contracts substantially mitigate the seeming deficiencies of single-settlement designs. In many jurisdictions the energy market conducted by the system operator is itself a residual market that accounts for a minor percentage of transactions. For example, the percentage is 5% in Victoria, 16% in Norway, and 8-10% in Sweden. Similarly, ISO-NE and Ontario intend to deduct bilateral contracts from the supply-demand configuration that is optimized under its dispatch protocols, so if their experience is like others then the percentage of transactions settled at the spot price will be minor. One can interpret this outcome as a privatization of the initial settlement; that is, it is conducted in the market for bilateral contracts that occurs before the residual market operated by the system operator.

Multiple-settlement designs bring their own problems. The most severe considerations concern physical feasibility and inter-market efficiency. Physical feasibility is jeopardized because an initial day-ahead market, in which suppliers are expected to internalize ramping constraints in their bids, is not guaranteed to be feasible, since there are few checks on their schedules to validate that operating constraints are met. This problem need not be severe if suppliers can offer portfolio bids for multiple generators, which increases the flexibility with which sale transactions can be fulfilled – and small suppliers can rely on long-term bilateral contracts to cover most of their output. Also, of course, the subsequent hour-ahead and real-time markets provide additional opportunities to adjust schedules to ensure physical feasibility.

The problem of intermarket efficiency is inherent in any design with multiple markets. Its main manifestations are in linkages among the energy, transmission, and ancillary services markets. However, even the linkages along a sequence of day-ahead, hour-ahead, and real-time markets present problems. Intermarket efficiency typically requires that each forward price is an unbiased estimator of the spot price, a statistical property associated with "rational expectations." In the first instance this depends on good information that provides unbiased predictions, but also it depends on sufficient arbitrage among the markets to eliminate persistent deviations. The key is contestability: it should be that a supplier who anticipates, say, a high spot price and therefore has an incentive to withhold capacity from the forward market in order to sell into the spot market, has ample opportunity and no impediment to doing so. This need not be the case if procedural rules and restrictions stand in the way. A simple example is the rule of ISO-NE that imports can be injected on short-notice only by a supplier included in the day-ahead schedule.

Summary

The fact that the marginal cost of energy is ultimately determined by the last generator in the merit order that can be incremented to meet the last MWh of demand on a spot basis is often viewed as sufficient reason to settle all transactions at the spot price, and besides the accounting requirements are simpler. This view ignores the role of intertemporal constraints, such as limited ramping rates, and the impact on revenue volatility, but most importantly it ignores effects on incentives. Designs that settle the transactions in each forward market at the prices that clear these markets ensure that suppliers are rewarded directly for the commitments contracted there. These designs are more complicated but they offer compensating advantages.

Multiple Markets and Inter-Market Efficiency

In its ideal form, an “optimized” pool manages everything, providing a single market for energy, transmission, ancillary services, etc. Using submitted data on availabilities, costs, and demands, and with complete data about transmission capacity, it establishes initial schedules and then supplements these based on developments in real-time. That is, it provides the services previously managed by vertically integrated utilities, or in some cases, established regional tight power pools. Here we address some of the issues that arise when this unified market is replaced by multiple markets of one form or another. We assume that transmission scheduling and real-time system control is conducted by a system operator (SO) who can drawn on ancillary services and supplemental energy offers to maintain system security, balancing, and load following. We divide the discussion between parallel markets and sequential markets.

Parallel Markets

Parallel markets exist elsewhere. One is NordPool in Scandinavia, which is a “marginal” market in the sense that less than 20% of energy is traded through the exchange. This structure, consisting of a large bilateral market for long-term contracts operating in parallel with a central market for spot trades, is common in various commodity industries – prominent examples are the metals markets, where as little as 5% of trades pass through the metal exchanges even though nearly all contract prices are pegged to the spot prices. The California design has made parallel markets a prominent issue. The debate between proponents of private bilateral markets and a pool was resolved there by allowing both. That is, in California the pool, called the Power Exchange (PX), is mandatory only for the incumbent utilities and only for a few years. Other private market makers called scheduling coordinators (SCs) can, like the PX and some large traders with direct access, submit balanced schedules for implementation by the SO (the California ISO). These private energy markets can operate in any format, as pools or bilateral contract markets or whatever they devise. The argument for the California design is that competition among alternative market designs is ultimately the best way to establish their relative merits. There are some practical reasons for establishing the PX as an official pool initially, and because the utilities are required to participate, it has a fair chance of establishing itself as the preferred market design.[53]

Efficiency could be jeopardized by different energy prices in the various markets. If the PX remains viable, this is unlikely in the long run, since non-utility traders can trade in any market with better prices, and in any case the non-PX market makers can themselves trade in the PX to erase persistent price differentials. Admittedly this argument is asymmetric, because as a pure market-clearing mechanism the PX cannot trade in other SCs' markets. The problem could be more substantial in the short run, since on any particular day the energy prices in the various markets might differ. The solution adopted in California is to allow inter-SC trades of adjustment bids, and in the real-time market, incs and decs that need not be paired within the same SC, and indeed for load following need not be paired at all.

The long term problem is the viability of the PX. Its role as an official market that assures open access, uniform pricing, and transparent operations would presumably not be filled by private markets. Its survival in competition with other SCs is jeopardized by its charter restriction to market clearing. For example, it cannot trade for its own account with other SCs (nor <u>in</u> their markets, although they can trade <u>in</u> the PX) to arbitrage the markets for energy and transmission. Another consideration stems from regulatory concerns. An official exchange or pool is easier to monitor and regulate. And if the market-making function for a critical commodity like electricity were dominated by private interests then new regulatory authority might be required to intervene in these markets to assure service in the public interest. This scenario has not occurred in the other basic commodity and service industries that have been deregulated, so it must rely on some aspect peculiar to the electricity industry. The presumed candidate is a market maker so successful that it can capture monopoly rents, but our impression is that the authority of electricity industry regulators is so pervasive as to make these concerns moot at present.

Inter-Market Efficiency

A pool tries to eliminate inefficiencies by a centralized explicit optimization based on submitted cost and engineering data, some of which is monitored for accuracy. The program allocates quantities subject to system constraints, but it also obtains shadow prices used for settlements. In principle, a dual formulation could be implemented as a single market with explicit prices determined by simultaneous clearing of the markets for each of the main ingredients, such as energy, transmission, and ancillary services. Several designs have been proposed for conducting these markets simultaneously, and at least one has received some experimental testing. For example, in one version the system

---

[53] The practical reasons include monitoring of the market power of incumbent utilities, and using the PX price to settle long-term contracts with what in the U.S. are called qualified facilities (QFs) under the 1978 PURPA regulations.

operator (SO) continually monitors transactions in a bilateral market based on posted bid and ask prices for energy, and then using the energy flows implied by these transactions, the SO solves a simplified dual problem that imputes shadow prices for injections at each node.

In practice, however, these markets are usually conducted in a sequence reflecting the fact that transmission demand is derived from energy transactions, and the supply is fixed. Similarly, the demand for ancillary services is nearly proportional to the demand for energy, since most system operators maintain reserves on that basis, and the supply consists mostly of residual generation capacity after accounting for the main energy transactions. Thus, the typical structure is a cascade in which the initial market is for energy, followed by a transmission market in which energy flows are adjusted to keep within the transfer capacity, then a market for ancillary services such as spinning and non-spinning reserves (for which some transfer capacity was previously set aside). These forward markets on a day-ahead (and perhaps hour-ahead) basis are followed by a real-time market in which the SO draws on supplementary offers to maintain system balancing on a short time scale, and when these are insufficient or expensive, calls on the ancillary services held in reserve.

The sequential market structure is convenient administratively and potentially as efficient as a simultaneous market. Realization of this potential depends, however, on several factors. The most obvious requirement is that the clearing prices must be tightly linked:

- The forward price for energy should be an unbiased estimator of the subsequent spot price.
- Traders transacting in the energy market should have accurate expectations about the usage charge that will be imposed later for transmission.
- Sales in the energy market should be based on accurate expectations about the opportunity cost of committing capacity there as opposed to offering it as an incremental bid in the transmission market or as reserve capacity in the ancillary services market.

The key to all three of these requirements is the accuracy, or at least the unbiasedness, of expectations about subsequent prices. Power markets are generally considered good candidates in this respect because they are repeated daily, basic energy and transmission capacity is largely fixed in the short term, and aggregate hourly demand can usually be estimated a day ahead within a few percent points – although unplanned outages and extreme weather conditions can produce larger discrepancies occasionally. In addition, that part of stochastic price variation that is insurable can be hedged via financial contracts, such as TCCs and CFDs.

Nevertheless, these favorable characteristics must be complemented with design features that provide structural support for the formation of accurate expectations. The most important is that all markets in the sequence must be easily contestable so that any significant price differences can be erased by arbitrage. Thus, systematically high prices for ancillary services should induce higher supply bids in the energy market from suppliers who recognize that they could leave some capacity uncommitted there in order to offer it as spinning reserve. And, systematically high usage charges for transmission should attract ample incremental and decremental bids that enable the SO to reduce congestion cheaply. The most important requisite for contestability is that participation in each market is voluntary, so that traders can move from one market to another to exploit apparent price advantages.

The problem lies in the term "systematically" above, since on any particular day it could be that higher or lower prices in subsequent markets were not anticipated in earlier markets, especially the energy market. Some of these unanticipated discrepancies can be reduced by provision of informative data and predictions by the SO and by market makers; e.g., the manager of the energy market can provide reports on inter-zonal imbalances after each iteration or bilateral transaction in the energy market so that traders can better estimate the magnitude of the inter-zonal balancing that must be solved in the subsequent transmission market.

A useful structural mechanism provides corrective markets that take account of the discrepancies. The following provide some indication of how this is done in the California design.

- One example is the provision for both day-ahead markets and a repetition (typically on a smaller scale) in hour-ahead markets (actually, two hours). Thus disparities detected after the close of the day-ahead markets encourage trading in the hour-ahead markets to exploit the price differences.
- Another is that after the initial calculation of day-ahead usage charges by the SO the non-PX scheduling coordinators are allowed to trade adjustment bids before submission of their final schedules. Also, the non-PX scheduling coordinators can trade in the PX in order to arbitrage price differences between their markets.
- A third is that portfolio bids are allowed in the day-ahead energy market, so that commitments of individual generation units need not be specified until after the hourly clearing prices for energy and the interzonal power flows are established.
- A fourth is that the day-ahead energy market is conducted iteratively, which allows traders to develop some consensus about the likely pattern of energy prices across the hours of the next day, which in turn reflect expectations about transmission, ancillary services, and real-time prices.
- Lastly, the ancillary services markets are also conducted in a cascade, so that bids rejected for one service, say spinning reserve, can be carried over to compete for another service, such as non-spinning reserve.

Despite these provisions, the link between the energy and transmission markets remains the most vulnerable. An extreme occurs when the adjustment bids, if they are voluntary, are insufficient to clear the market for transmission, but more routinely it could be that usage charges are too volatile to enable reliable predictions by traders in the energy markets. Transmission pricing based solely on congestion is inherently volatile because the usage charge across an interface can be zero if capacity slightly exceeds demand, and significantly positive if the unadjusted demand slightly exceeds capacity. And other minor procedural aspects can impair predictability; e.g., if multi-zone portfolio bids are allowed then the power exchange cannot provide reliable estimates about the magnitude of the interzonal flows implied by the tentative trades during the iterative process; and prohibition against trading adjustment bids among scheduling coordinators (adopted in

California as a "simplification" for the first few months to facilitate startup) can yield exaggerated usage charges because an increment from one SC cannot be matched with a decrement from another.[54] For these reasons it is clear that a design priority is to strengthen the link between the day-ahead energy and transmission markets, and perhaps to adopt a design that integrates these two key markets.

Limitations of Time-Differentiated Prices

Essentially all jurisdictions with competitive wholesale markets for electricity establish prices according to the time of day. Typically these prices apply to intervals of an hour (California), half-hour (U.K.), or as short as five minutes (New England). The shorter intervals are used for real-time balancing markets, the longer intervals for day-ahead markets, and some markets for long-term bilateral contracts divide the day into only two intervals corresponding to the peak and off-peak periods. Time-of-day pricing in competitive markets corresponds to time-of-day tariffs in regulated markets.[55] Prices are also differentiated on other dimensions, such as location to take account of transmission constraints, but here we focus on the effects of time differentiation. To be concrete, we refer to the intervals for which prices are set as the hours of the day. In a fully decentralized system, this means that the markets are cleared for each hour, and in an centralized system with optimized dispatch, it means that the marginal cost of supply is calculated for each hour.

Electricity markets that rely on hourly prices are inherently limited in the productive efficiency that can be obtained. The source of inefficiency is the set of intertemporal constraints that limit the operations of thermal generators. The magnitude of the inefficiency is presumably small in a system whose main supply is generated by spilling stored water in reservoirs, since these sources can be turned on and off quickly. The intertemporal constraints are more significant in hydro systems in which spills through a cascade of reservoirs that must be coordinated, and in those dependent on the run of a river. The most significant constraints apply to thermal generators, especially large coal-fired plants, that are limited by the rates at which they can ramp up or down. Typically these rates are on the order of 1% of capacity per minute. Additional constraints apply to plants that are cycled each day, which incur costs warming up and cooling down their boilers.

To be specific, we focus here on ramping constraints that limit the flexibility of generators.[56] The effect of these constraints is to link each hour to the hours before and after, thereby invalidating the time-separability of costs that is necessary for time-

---

[54] The California design has inherent structural biases. The day-ahead transmission market relies on inc/dec pairs to balance interzonal flows, whereas the real-time market is not confined to matched pairs, and further, SCs pay the cost of interzonal balancing whereas the SO absorbs the cost of intrazonal balancing.

[55] In contrast, some regulated markets use so-called Wright tariffs that charge a retail customer, for each kilowatt of annual demand, a fixed fee plus a price per hour that the load uses that kilowatt. Wright tariffs mimic the long-term costs of electricity supply, which include the amortized investment cost of capacity and the hourly operating cost of a generator.

[56] In principle, similar considerations apply to demand-side applications, such as ovens, with limited flexibility.

differentiated pricing to be fully consistent with efficient operations. Various forms of nonlinear pricing could mitigate some of these effects, but in practice the payment systems are invariably linear: except for the payments for start-up and no-load costs allowed in some systems, the payment in each hour is the product of that hour's price and the supplier's metered quantity of energy provided. Bid formats do not include some relevant information, such as the longer-run costs of maintenance caused by ramping that stresses turbines and generators.

The problem can be seen most clearly by considering the operating protocol of a typical centralized system in which the system operator optimizes dispatch. Assume further that a single settlement is used: in each hour suppliers are paid for each MWh the marginal cost of energy ($/MWh), computed as the shadow price on the constraint that energy injections and extractions (plus losses) must balance. The system operator optimizes the system using a linear program (LP) in which, besides energy balancing, the constraints include many different operating limits, such as transmission capacities, and of course the ramping constraints of the various generators. A key feature is that this LP is solved repeatedly, as frequently as every five minutes, using a horizon looking forward that is typically 24 hours so that all the ramping constraints over the full daily cycle of the system are included.

The daily cycle can be envisioned as an initial solution of the LP at midnight, followed by updates every hour through the ensuing day. The initial solution at midnight relies on imperfect projections of supply and demand conditions, but it has the advantage of optimizing the operating levels of all generators in all hours. On the other hand, the re-calculation at noon has better information, but it must start from the generators' operating levels prevailing at noon, which may be sub-optimal under the currently prevailing conditions, and which may therefore be affected by additional ramping constraints that were not envisioned at midnight. Thus, the price calculated at noon reflects in part the costs of inaccurate projections at midnight, and might therefore be too high or low compared to an optimization based on perfectly accurate forecasts.

But even if with perfect forecasts, the hourly prices tend to be biased systematically. If a peak is predicted at noon then, based on the LP solution at midnight, some generators will be (slowly) ramped up beforehand to meet that peak and then (slowly) ramped down afterwards. The cost of the ramping before the peak is not recognized in the shadow price calculated at noon, since the LP takes the high level of the generators' current operating rates as a fortuitous initial condition for the next hour. Similarly, during the ramping down in the afternoon the high operating level of generators with limited downward ramping rates is taken as an unfortunate initial condition. The net result is that solving the LP repeatedly with a rolling horizon tends to depress the peak price, and raise prices in some other hours, compared to the full cost of serving the load over the full daily cycle.

It might be thought that this bias is inconsequential, because over the daily cycle the under-pricing in the peak hours is offset exactly by the over-pricing in some other hours. Indeed, in a purely cyclical system it is true that the imputed costs of the ramping constraints, expressed as shadow prices, average out to zero – what goes up must come down. But the incentive effects on suppliers are not so innocuous. The reason is the tension between those with flexible plants and those with inflexible plants constrained by ramping rates and the costs of equipment stresses. Depressed peak prices can be interpreted as disadvantaging those with flexible plants, who might otherwise have obtained higher profits from their generation in peak periods. For instance, they may see that at midnight they were scheduled to meet a portion of the peak load based on a full optimization of all units, taking account of the costly ramping constraints on less flexible plants, but when noon arrives the price they receive is depressed by the currently high operating rate of the inflexible plants achieved by ramping in earlier hours, with no accounting at noon for the prior cost of ramping up the inflexible plants.

The situation is somewhat different in fully decentralized systems in which each supplier self-schedules its units to meet its sales of energy into the market. In this case, a supplier must substantially internalize the effects of its intertemporal operating constraints, such as ramping limits and costs. If the day-ahead energy market proceeds iteratively to establish clearing prices for each hour that determine settlements, or if sales are made bilaterally via longer-term contracts, then a supplier can respond to the pattern of prevailing prices over the hours of the next day to select its entire operating schedule over the next daily cycle. Thus, it is paid in advance for the operating commitments it makes in advance, and for deviations in real time it is paid the real-time price. However, the restriction to hourly market clearings still limits the efficiency that can be obtained. This stems from the fact that typically the costs of a thermal unit stem mainly from the level and duration of an operating run over consecutive hours of operation, which have little directly to do with the time of day in which generation is provided.

Feasibility

Markets for electricity differ from those for other commodities because operations are more severely constrained by physical considerations. The few options for storing energy, such as pumping water into reservoirs, are expensive; most generators ramp slowly, roughly 1% per minute; and the energy balance and stability of the transmission system must be maintained continuously. The consequences of these physical considerations are evident in the increasingly stronger authority assigned to the system operator as the time of dispatch approaches. On the time scale of a few minutes the system operator has full control of scheduling, and on the shortest time scale of a few seconds the energy balance is maintained by automatic generation control (AGC).
The latitude for competitive markets to allocate resources in a decentralized fashion therefore increases in proportion to the time before dispatch. Privately managed markets suffice for trading long-term bilateral contracts. Exchanges for clearing markets for clearing day-ahead and hour-ahead trades of forward contracts are more centralized. Usually the system operator manages reserves itself or conducts the auction markets in which it acquires ancillary services, and it conducts the real-time market for the spot trades used for load following and balancing. Although the generation dispatched in real-time by the system operator is usually small, on the order of 3%, it is critical to system security – and in some designs the spot price for balancing energy is used to settle all previous transactions.

A central issue in market design is the extent to which the role of the system operator (SO) can be minimized, and correspondingly the role of market processes can be increased. In part this is a technical issue and one that moves steadily with advances in technology and computer software. For instance, proposals for re-designing the U.K. system envision reducing the SO's window of control from a day ahead to four hours before dispatch. And of course the speed, flexibility, and reliability required of command and control systems increase proportionately with the volatility of demand and the complexity of the supply and transmission system.

The other part of this issue, however, is the extent to which decentralized markets can be relied on to produce feasible schedules. The problem occurs in different forms in the ancillary services, transmission, and energy markets, which are addressed in sequence below. Because physical feasibility is most important in highly decentralized systems like California, we use it as the basis for the discussion. Much more than most systems, California relies on voluntary bids offered in separate markets, rather than a centralized allocation of resources, so it presents the issue in its extreme form.

Ancillary Services

The SO's minimal requirements for ancillary services are usually set by technical standards established by the industry. For instance, California adheres to the standards of the Western States Coordinating Council (WSCC), which requires 7% ready reserves of supplemental energy, of which for thermal generators 3.5% must be spinning reserve that can be ramped within 10 minutes and maintained for two hours.[57] The SO acquires these reserves a day ahead via a sequence of auctions for the various kinds of reserves, such as regulation, spinning, non-spinning, replacement, black start, etc. The typical situation is that ample supplies are offered for spinning and other energy reserves. This is a natural consequence of the preceding energy market, since generators with residual capacity after the close of the energy market can obtain further revenue from bidding into the reserve markets. Regulation is more problematic, however. Not all units are equipped to provide AGC, and further the payments for generated energy tend to be small, since each AGC unit is required by WSCC standards to return to its set point every ten minutes. Thus, a supplier's revenue from AGC is obtained almost entirely from the payment for reserved capacity, measured in terms of the unit's range of incremental and decremental generation.

(Previously, a complicating factor was that FERC set limits on the capacity prices payable to units owned by the former utilities, which discouraged them from offering bids for AGC.) A complicating factor is that in the absence of sufficient voluntary bids, the SO must rely on strategically located "regulatory must-run" (RMRs) units for which long-term contracts have been established previously, under terms mandated by FERC. These contracts require very high payments (typically about twice the market price) whenever a unit is called, and the cost is inflated further by the fact that in order to

---

[57] We ignore here those potential infeasibilities peculiar to ancillary services that are addressed via monitoring. For instance, reserved units are tested occasionally to ensure that they can in fact meet the requisite start-up and ramping requirements, and penalties are imposed for inadequate performance.

obtain, say, a 100 MW range of AGC regulation service the SO must order a set point of, say, 200 MW so that lower limit of the unit's range of operation does not fall below its feasible minimum level of generation. The net effect of the FERC limits on capacity payments is to increase the cost of ensuring sufficient regulation via AGC. In practice, ad hoc solutions have been devised, such as indirect payments for capacity. The difficulties in California of obtaining sufficient low-cost supplies of AGC stem ultimately from regulatory intervention that prevents market clearing on the basis of voluntary bids. These effects of price caps are indicative of the costly consequences of ensuring physical feasibility whenever interventions or restrictive rules prevent normal clearing of the markets. [[Add here the heavy use of AGC for load following in shoulder hours.]]

Transmission

California is virtually unique in relying on voluntary adjustment bids to alleviate congestion in the day-ahead inter-zonal market for transmission. Several other jurisdictions, such as NordPool, ensure sufficient adjustments by drawing upon the bids submitted in the day-ahead energy market. Thus, a bid accepted for 200 MW in the energy market might be adjusted up to 250 MW or down to 150 MW to alleviate congestion. Typically a unit is incremented if it is in an importing zone or decremented if it is in an exporting zone. Because adjustment bids are voluntary and distinct from suppliers' energy bids, however, California's adjustment market occasionally failed to clear in the early months of operations, although only in cases in which importers into California provided insufficient decremental bids. The consequences for the SO are potentially severe, since inter-zonal congestion that is not alleviated day-ahead might have to be eliminated in the real-time market, which was designed only for intra-zonal congestion management on a much smaller scale. The cost of inter-zonal congestion management in real-time can be much higher; moreover this cost is absorbed by the SO and ultimately paid by all participants rather than only those causing the congestion. The expedient solution adopted by the SO imposes a default usage charge, currently allowed as high as $250/MWh, for transfers across an interface for which the adjustment market fails. This creates a strong incentive for submission of adjustment bids, so strong in fact that the CA-PX thereupon mandated adjustment bids from each of its participants for the full range of each unit's capacity. The default usage charge results in novel outcomes, the most important being negative prices; e.g., if there is no congestion except on an import interface with Arizona for which insufficient decremental adjustments were offered, then a $20/MWh energy price within California implies an energy price net of the default usage charge that is $–230 in the Arizona zone for suppliers importing from there into California. It takes just one experience paying a punitive price for generation to motivate a supplier to ensure that in the future its energy bids are accompanied by adjustment bids.

The lesson from California's adjustment market is that reliance on voluntary markets must be tempered with recognition of the overriding importance of ensuring physical feasibility of real-time dispatch. Measures such as a default usage charge are apparently necessary to ensure that participants are aware of the severe consequences of infeasible bids, and are motivated therefore to provide sufficient adjustment bids to enable congestion to be alleviated. Even though a default usage charge may seem onerous, it is not punitive to the extent that it reflects the expected cost to the SO of alleviating large-scale inter-zonal congestion during real-time operations.

A proviso to this lesson is that the eventual form and role of adjustment markets is not certain. Congestion pricing of transmission that assigns an injection charge at each bus (nodal pricing) produces substantial volatility and is ill-suited to secondary markets for point-to-point TCCs, or futures contracts such as those traded on the NYMEX commodities exchange. Pricing on the basis of hubs or zones reduces volatility and supports liquid markets for firm transmission rights (FTRs) such as TCCs with or without scheduling priorities. Because FERC's Order 888 nearly mandates provision of financial instruments for "price certainty" via some form of FTRs, and its policy is to encourage issuance of FTRs for 100% of inter-zonal capacity, it is likely that the adjustment market conducted on a day-ahead basis by the SO will eventually be replaced by secondary markets for trading FTRs in those jurisdictions allowing decentralized markets.

Energy

The most complicated issue of physical feasibility concerns suppliers' schedules submitted for dispatch after the close of the energy markets, including both the markets for long-term bilateral contracts and those such as the PX that clear markets on a day-ahead and hour-ahead basis.

In centralized systems (such as those in Alberta, New England, New York, PJM, and the U.K.'s E&W Pool prior to 1999) the SO converts bids into optimized day-ahead operating schedules that are continually updated, and participants are required, subject to penalties, to follow dispatch instructions. In contrast, a highly decentralized system such as California is based on the principle of self-scheduling. That is, each participant retains responsibility for constructing and submitting a feasible operating schedule that fulfills its transactions concluded in the energy market. Deviations from a day-ahead schedule can be corrected in the hour-ahead market, or corrected by the SO in real-time using resources from its supplemental energy market – and each deviation is charged or paid the price in the subsequent market. Unlike a centralized market in which the SO ensures that operating constraints are met, such as limits on ramping rates, a decentralized market with self-scheduling imposes only a few consistency checks and relies mainly on participants to ensure the physical feasibility of their schedules. It may not be evident to the SO whether a seeming infeasibility in a submitted day-ahead schedule reflects an error by the participant, or an intention to achieve feasibility by trades in the subsequent hour-ahead or real-time markets. It appears that the SO's inability to ensure feasibility or issue dispatch instructions might lead to higher reserve requirements. (For example, in California the SO's inability to ensure sufficient regulation via AGC has led it to mandate that the PX include "regulatory must run" units in its schedule when so instructed.)

Nevertheless, the main problem of feasibility lies with suppliers, who bear direct responsibility for feasibility via self-scheduling of their unit commitments. In practice this has two components. One component is financial, in the sense that a supplier absorbs the revenue consequences of schedules that are unprofitable due to start-up and no-load costs and ramping constraints. In California, financial considerations were the main motive for the proposed implementation of an iterative format for the day-ahead energy market. The argument was that iterations of the market clearing process would enable suppliers to see the pattern of hourly prices over the day and thereby respond optimally. The iterative format has not yet been implemented because in practice several other options enable sufficient protection against unprofitable schedules. These options include portfolio bidding in the day-ahead market that provides additional scheduling flexibility, and additional trades in the hour-ahead and real-time markets; further, a small supplier has ample options for long-term bilateral forward contracts that cover most of a unit's capacity.

The second component is physical feasibility. In the California PX the single-iteration format runs the risk that the portion of a supplier's bids accepted at the clearing prices is physically infeasible due to ramping constraints. For instance, it can be that a supplier sells a large quantity in hours 10 and 12 and none in hour 11, when in fact its unit cannot be ramped down and then up within the single hour 11. In this case, the supplier is sure to deviate from its accepted schedule, the likely outcome being that it sells additional energy in the hour-ahead market for hour 11, or failing that, in the real-time market – possibly accompanied by purchases in hours in 10 and 12 that reduce the ramping required in the hour between. One can interpret this as a financial risk, well within normal commercial practice, and one that is not severe if proper attention is paid to predicting prices accurately based on the daily repetition of the market and good information about demand and supply conditions; e.g., a rule of thumb is that the demand load in each hour is usually predictable within 3% on a day-ahead basis. From the viewpoint of the SO, however, these financial risks translate into substantial uncertainty about how the submitted day-ahead schedules will translate into real-time operations. Inevitably they limit the SO's ability to predict and prepare for contingencies in real-time, and increase the burden on the real-time market to cope with imbalances that are not resolved until shortly before dispatch. The ultimate solution is a vigorous real-time market that provides the SO with ample resources to remedy imbalances. This has generally been the case in California, but the anxiety persists that some event will arise in which imbalances of substantial magnitude will need to be solved in real time without adequate resources available.

Although these problems of physical feasibility are most prominent in highly decentralized systems, even highly centralized ones are not immune. For instance ISO-NE allows imports and exports on short notice that could obviate all the advance scheduling that was undertaken before real-time dispatch. Similarly, Alberta allows short notice of which among several submitted schedules an importer or exporter will use, and it allows other suppliers to re-declare their schedules as close as four hours before dispatch. The central fact is that in any system the intertemporal ramping constraints limit the physical flexibility of the system response, while the incentives of traders in the market encourage maximum exploitation of financial opportunities on short notice. Imports and exports are particularly sensitive, but in principle the same tension between physical constraints and financial motives affects all operations. Traders in the market would like the SO to have infinite flexibility, but this is physically impossible, so necessarily the SO must enforce limits on the response rates that its participants can impose. Centralized systems provide the SO with greater control of dispatch (and all systems authorize full control in emergencies) so that anxieties about physical feasibility are minimized, but always there is pressure from suppliers to allow exceptions that would enable them to exploit financial opportunities quickly, on the shortest time frame that SO will allow.

The risks of physical infeasibility set one of the lower limits on how far the role of the SO can be reduced. The so-called min-SO, the system operator with minimal responsibility for only real-time load-following and regulation, is also the one facing the greatest risk of physical infeasibility, and therefore the one most likely to exercise heavy-handed intervention to prevent instability or failures in the transmission system.

Mitigation of Market Power

In most jurisdictions, restructuring the electricity industry to establish competitive markets begins by reversing the vertical integration of the incumbent utilities. In some such as the U.K. the utilities were government owned enterprises reorganized as private corporations whose shares are sold to the public. In others the utilities are already privately owned but were regulated by government agencies. Previously the "regulatory compact" provided monopoly franchises for their service areas in exchange for an assured return on shareholders' investments. In both cases the basic model for deregulation was described by Paul Joskow and Richard Schmalensee in *Markets for Power*, MIT Press, 1983. They envisioned an industry in which the energy and "wires" parts of the business are separated. The energy part is not regulated whereas the wires part remains lightly regulated, based primarily on principles of common carriage to provide open access on nondiscriminatory terms. The production and marketing of energy is organized around competitive wholesale markets for electricity. On the wires side, use of transmission assets is allocated by a system operator while retail distribution remains a regulated monopoly.[58] These features require that the energy generation and marketing businesses are separated from the wires businesses that own transmission and distribution assets.

The key step in restructuring is to establish the conditions for competitive energy markets – and also transmission and ancillary services if they are allocated via markets. The usual view is that if the supply side of the energy markets is competitive then ease of entry into retail marketing assures efficiency of the markets overall. The central issue is therefore to establish the conditions for competitive supply. Several approaches have been taken.

Divestiture. The direct approach is to require or encourage utilities to divest a substantial proportion of their generation plants. This has been the policy in all those states in the U.S. that have restructured – except Montana where Montana Power volunteered. For example, in California the two largest utilities sold all their thermal

[58] In the U.S. these principles for regulation of interstate transmission are codified in the Energy Policy Act of 1992 and in Orders 888 and 889 of 1994 by the Federal Energy Regulatory Commission. However, in some states, municipal utilities have the option to retain control of their transmission assets and entitlements. In states such as California this option has been exercised, with the result that the system operator manages only a portion of the transmission within its control area.

plants fired by fossil fuels, and given the large number of independent power producers in the state and substantial supplies of imports this was deemed adequate to prevent large sustained influences on market prices. In contrast, the U.K. initially established just two private corporations with large shares of the fossil-fueled capacity that in retrospect have been vigorously criticized as retaining significant abilities to affect prices.[59] And in Ontario explicitly and Alberta implicitly the provincial government has committed itself not to require divestiture.

Much of the initial resistance to divestiture stemmed from anxiety about stranded costs, for which ratepayers would be liable entirely or in substantial part. In fact, plants sales have been at prices above book value so there have been no stranded costs.

On the other hand, divestiture has not improved substantially the competitiveness of energy markets. This is due to "full-fleet" sales; e.g., Duke Energy bought PG&E's plants in California, while PG&E's unregulated Energy Services subsidiary bought the plants of New England Electric System [?]; and in New York the plants of NYSEG were bought by [?]. In New England the net result is that two firms, NU and PG&EES, own nearly half [? – see Cramton-Wilson report] the capacity and through partial ownership positions are potentially privy to bids from over 60% [?] of the capacity.[60] Invariably the sales have been conducted privately by investment banks engaged by the seller (not the ratepayers' consumer advocacy branch of the state regulator) rather than open auctions; this lack of transparency that would guard against self-dealing, side-deals, and affiliate favoritism could become a problem in subsequent divestitures if current practices continue.

Contract Cover. An indirect approach is to provide incumbent utilities with supply contracts, sometimes called legislated hedges, that fix the price of delivered energy for a substantial fraction of their capacities. This was the policy in the U.K. and Alberta for the first few years. The fraction of a utility's capacity that is covered and the guaranteed price can in principle be set so that the utility's incentives are to act like a competitive producer in the markets where it sells its residual output.[61] This is usually seen as an interim solution; e.g., in Alberta the hedged contracts between the major utilities' generation and distribution subsidiaries effectively precluded entry by new firms. Also, if the contract cover is too large or small then the resulting market prices can be too low or too high.

Auctions of Biddable Entitlements. An innovative approach considered in New York and adopted in Alberta obligates the incumbent utilities to sell at auction long-term entitlements to a portion of their output. This is primarily a device to enable marketers to acquire energy supplies that they can resell retail, but because these supplies can also be bid into the wholesale market, they exert competitive pressure on incumbents. Of course

---

[59] The U.K. experience is summarized by M. Armstrong, ? Cowan, and J. Vickers, ??, MIT Press, 19??, and analyzed quantitatively by F. Wolak, "…" 199? and Catherine Wolfram, "…", 199?.
[60] NU's plans for divestitures have not been revealed.
[61] See Richard Green, "…," … for a detailed analysis of the incentive effects and the calculation of the right fraction of contract cover and the hedged price.

the entitlement contracts must be carefully designed to preclude tactics by incumbents that disadvantage competitors.

Contestability via Entry. Absent divestiture, one long-term solution is to encourage entry of new producers. For example, in the U.K. new entrants provide over 10% of supplies. Many jurisdictions provide subsidies to renewable sources (wind, etc.), and some marketers offer supplies from renewable sources at higher prices, that encourage entry. Co-generation facilities are usually accorded favorable treatment and account for a steady increase in supply. The dominant factor, nevertheless, is usually the construction of new combined-cycle gas turbines (CCGTs). These have the advantages of low fuel costs, high thermal efficiency, low emissions, small scale, and easy siting. Their rate of entry can, however, be impeded by a large stock of existing capacity using coal and nuclear fuels that may take many years to be mothballed or decommissioned.

Contestability via Imports. Often neglected is the opportunity to increase contestability by increasing the transmission capacity for imports. For example, in a province such as Ontario, where divestiture has been ruled out politically, imports from the U.S. could diminish the market power of Ontario Hydro, which owns over 85% of capacity. Dominant incumbents can hold considerable sway over the governing board that decides on transmission expansions, however, and naturally they are opposed to any increase in import capacity.

Threat of Re-Regulation. New Zealand is apparently unique in using an explicit threat of re-regulation as the main constraint on the exercise of market power by the dominant firms. However, the regulator in the U.K. has also indicated its intention of intervening if average prices are deemed excessive.

All these approaches to market power mitigation are strongly affected by current political considerations as well as historical residue from the previous era of regulation. Large states in the U.S. such as California and New York are endowed with significant supplies from independent power producers and from imports, so a mild program of divestiture sufficed to establish the primary conditions for competitive energy markets. (Pennsylvania-New Jersey-Maryland (PJM) has yet to convince FERC that it can sustain competitive markets, but this may have more to do with reliance on nodal pricing that conveys substantial market power to incumbents in critical locations.) In contrast, the provinces of Alberta and Ontario in Canada excluded divestiture on political grounds, so explicit measures are required to mitigate the market power of the dominant incumbents.

All jurisdictions maintain some semblance of market power surveillance, and indeed FERC has required formal surveillance in the U.S. jurisdictions. However, the prospects for success are not bright in those that have undertaken little or no divestiture. In the U.K. the struggle between the two major firms and the regulator has been ongoing (resulting in further divestiture of plants), and in Alberta the committee charged with market power surveillance disbanded, complaining that it was powerless to deter abuses.

The future may present the new challenge of mitigating the influence of dominant market makers. The scheme proposed by Joskow and Schmalensee in 1983 relied on a common-carrier transmission system, and is presently implemented in most jurisdictions by assigning authority for transmission management to a system operator. Organized as a public-benefit corporation, the system operator allocates transmission capacity on a forward basis to ensure security and manages regulation and congestion in real time using resources offered in a spot market and reserves obtained from prior purchases of ancillary services. From the perspective of suppliers and marketers, however, the charges imposed by the system operator to reflect the costs of alleviating congestion, either forward or spot, produce a price risk that they prefer to avoid by contracting ahead for both energy and transmission. Some jurisdictions avoid this price risk by using postage-stamp usage charges and including the costs in uplift charges, but in those that rely on congestion charges the solution has been to issue firm transmission rights (FTRs, which are essentially required in the U.S. by FERC Order 888 assuring options for "price certainty"), usually in the form of a refund of the usage charge but also including scheduling priority whenever congestion cannot be fully eliminated using the adjustment bids submitted. The main consequence of FTRs is essentially to privatize the transmission assets by transferring the forward market for congestion management into the secondary market for FTRs. The availability of FTRs enables dealers in the energy markets to offer forward contracts that bundle energy and transmission. In time, it could develop that a single private market for energy *cum* transmission acquires a dominant position, a development that was originally thought to be precluded by the monopoly authority of the system operator to manage the transmission system. In this case, surveillance and mitigation of the power of the market maker will be as important as was originally intended for the major suppliers.

Investments in Transmission Capacity

The eventual consequences of restructuring wholesale electricity markets are largely determined by the investments made in later years. These investments depend heavily on the incentives created by the daily markets. In some states the motive for restructuring is to avoid repeating a pattern of utility investments viewed in hindsight as uneconomic, so the benefits and costs will ultimately be measured partly by the effects on private investment incentives. Like most economists we trust that competitive markets for generation will provide correct incentives for investments, especially now that the minimum efficient scale of combined-cycle generators is rather small (approx. 200 MW) and their siting requirements are moderate. Here we describe briefly a few aspects of this issue in the narrow context of transmission, where trust in the market is not justified.

The central unsolved problem in competitive power markets is how to ensure efficient investments in new or expanded transmission capacity. Because the difficulties of acquiring of new rights-of-way are often cited as the chief impediments to new transmission lines, we emphasize that the problem of investment incentives is already evident when we consider only expansion or strengthening of existing lines, so this is our focus.

The basic fact is that a transmission system is essentially a public good, that is, an instrument of public convenience or necessity as customarily defined in legislation. Efficiency can therefore be defined only in terms of the benefits to users collectively. There is no general theory of how to encourage efficient investment in transmission via decentralized markets, because the pervasive externalities and scale economies ensure that privately owned lines cannot capture in revenue the entirety of the collective benefit. For instance, congestion pricing captures a portion when a line is congested but none when it is not, and access fees and volumetric rates cause well-known departures from efficiency. There will be cases in which congestion charges are sufficiently high to justify private investments in capacity expansion to reduce these charges, but these alone will be insufficient. A few jurisdictions retain for government agencies the authority to order capacity expansion in the public interest (one is the Ontario Electricity Board), and in others such as Australia and the U.K. the grid operator is also the owner and is assured a regulated rate of return for prudent investments in the public interest. Some state commissions are proceeding similarly by continuing to assure a regulated return on utility-owned transmission assets via surcharges on energy.

The problem, however, is that the chief player in decisions about capacity expansion is the system operator, mainly because it is the repository of the relevant information and expertise. These decisions therefore depend crucially on the outcome of debates in the governing body of the system operator. The stakeholder interests represented on a governing board differ fundamentally on transmission issues. Expanded transmission can mean higher prices for exports or lower prices locally due to new competition from imports. Demanders in low-cost areas can see the regional price equalization resulting from transmission expansion as a subsidy to local suppliers that comes from their pockets – and reversely for suppliers in high-cost areas, who are equally opposed to transmission expansion. It is not clear that these conflicting interests about distributional effects will be resolved in the collective interest.

An important exception pertains to mitigation of market power. If transmission capacity is sufficient to allow ample imports then the market power of local suppliers is eroded. Even short lines can have large effects, as in the case of the reliability-must-run units needed for local reliability in California. Because FERC rules prevent domination of governing boards by major suppliers, there is substantial likelihood that capacity expansion to improve the contestability of markets by imports will succeed. It is uncertain whether these rules work equally well to prevent domination by large demanders and consumer groups interested in large but possibly uneconomic expansion of transmission capacity, especially if these demanders are assured a regulated return on their transmission assets.

Concluding Remarks

Our examination of the architecture of wholesale electricity markets presumes that the ingredients for effective competition are present. It is important to emphasize further that market architecture is distinctly secondary in importance to market structure, in the sense of competitiveness or contestability. Monopoly power in generation, or local monopolies due to transmission constraints, can impair efficiency regardless of the market design implemented. Oligopolies are inherently more damaging to the public interest in power markets because their daily interaction offers ample opportunities for punishment strategies to police collusive arrangements, whether explicit or implicit. Thus, structural solutions to the market power of dominant incumbents are necessary. In the same way, procedural rules are less important than architecture: no amount of fiddling with procedural rules can overcome major deficiencies in the links among the energy, transmission, and ancillary services markets. There is therefore a natural priority in the design process that starts with ensuring a competitive market structure, proceeds to the selection of the main market forums, and then concludes with the detailed issues of governance and procedures. Some procedural rules, of course, must be designed to mitigate market power and prevent collusion; e.g., it is usual to maintain the secrecy of submitted bids to thwart efforts by a collusive coalition to punish deviants.

An aspect omitted here is the role of transaction costs. This consideration affects all three stages of the design process. Procedural rules must obviously be designed to avoid unnecessary transaction costs, but it is well to realize too that a complex array of decentralized markets imposes burdens on traders, who may well prefer a simpler structure that avoids managing a complex portfolio of contracts, bids, and schedules. A simple design can also promote competition by bringing all traders together in a few markets with standardized contracts, bid formats, and trading procedures. The virtues of simplicity can be especially important in jurisdictions with few participants and small volumes of trade.

Appendix A: Glossary of Acronyms

AC	Alternating Current, the standard system in North America
AGC	Automatic Generation Control, used for instantaneous load following
ANG	Australian National Grid, the consortium formed from the state systems in South Australia, Victoria, and New South Wales, later to include Queensland. Operations began in 1999.
APX	Automatic Power Exchange, operating in California and New York
AS	Ancillary Services, sometimes written as A/S
BEEP	Balancing Energy Equilibrium Program(?). In the California ISO the BEEP stack is the collection of offered incs and decs in merit order
BPA	Bonneville Power Administration, a federal agency with hydro and transmission resources in Washington and Oregon
CA	California
CCGT	Combined-Cycle Gas-Fired Turbine
CEC	California Energy Commission
CFD	Contract for Differences, an energy contract that includes a price hedge
CFTC	Commodity Futures Trading Commission, the federal agency that regulates futures markets
COB	California-Oregon Border, a delivery point for NYMEX futures contracts
CT	Combustion Turbine
CTC	Competitive Transition Charge, in California mainly to pay residual debt obligations of utilities for investments prior to restructuring.
DC	Direct Current, the standard system in Europe, but used in the U.S. for some major transmission lines, typically not integrated with the AC grid
DOE	Department of Energy, a federal agency in the U.S.
DPU	Department of Public Utilities, in states such as Massachusetts
DTI	Department of Transportation and Industry, in the U.K.
EBB	Electronic Bulletin Board, a means of posting information or bids
EEI	Edison Electric Institute, an industry research organization in the U.S.
EOB	Electricity Oversight Board, an agency created by statute in California
EPRI	Electric Power Research Institute, an industry research organization
ERGOT	Texas system operator, or its control area
E-W	England-Wales, the jurisdictional area of the principal U.K. system operator
FAF	Fuel Adjustment Factor, included in tariff specifications to account for variations in fuel costs
FERC	Federal Energy Regulatory Commission, the federal agency in the U.S. empowered to regulate inter-state trade and transmission of electricity, gas, and other energy sources
?Fin-El	The market for financial instruments used as hedges against spot prices of electricity in Finland
FTR	Firm Transmission Right, a financial hedge against transmission usage charges that includes a scheduling priority

GMM	Generation Meter Multiplier, used to approximate transmission losses
GOAL	? Generation Optimization and Allocation….?, optimization program used in the England-Wales system
HOL	High Operating Limit, the maximum production rate of a generator
IOU	Investor-Owned Utility, as opposed to a municipal utility or other public enterprise or agency
IPP	Independent Power Producer
ISO	Independent System Operator
ISO-CA	California ISO
ISO-NE	New England ISO
ISO-NY	New York ISO
LDC	Local Distribution Company, a synonym for UDC.
LMP	Locational Marginal Pricing, a synonym for nodal pricing.
LOL	Lower Operating Limit, referring to the minimum stable operating rate of a generator
LP	Linear Programming, a method of formulation and calculation used for optimization of unit commitments and schedules
kW	Kilo-Watt, a measure of power, energy per unit time
kWh	Kilo-Watt-hour, a measure of energy
MC	Marginal Cost
MCP	Market Clearing Price
MEE?	Mercado Elec… de Espana, the energy market in Spain
MERC	Chicago Mercantile Exchange
MSC	Market Surveillance Committee, in California a monitoring body, in New Zealand a judicial body.
MSG	Minimum Scale of Generation, in Australia – same as LOL.
MW	Mega-Watt, measure of power, energy per unit time
MWh	Mega-Watt-hour, a measure of energy
NE	New England
NEPool	New England Power Pool
NordPool	Scandinavian system operator for Norway and Sweden, later to include others such as Denmark and Finland
NERC	North American Electric Reliability Council, an organization delegated responsibility by FERC for reliability standards and procedures
NRRI	National Regulatory Research Institute, a research organization of the state regulatory agencies
NY	New York
NYMEX	New York Mercantile Exchange, an exchange for financial contracts such as futures and for trading commodities
NZ	New Zealand
OASIS	?Open Access Status(?) Information System, mandated by FERC to be maintained by major entities, such as transmission owners and operators, under its jurisdiction.
Offer	Office of Energy Regulation, in the U.K.
O&M	Operationing and Maintenance costs
OPF	Optimal Power Flow, a computational procedure

PJM	Pennsylvania-New Jersey-Maryland
PPA	Power Purchase Agreement
PSA	Power Supply Agreement
PSC	Public Service Commission, in states such as New York
PUC	Public Utility Commission, in states such as California and Oregon
PURPA	Public Utilities Regulatory Policies Act, 1978
PX	Power Exchange, in jurisdictions such as California
QF	Qualified Facility, as specified in PURPA
RMR	Reliability Must-Run, a plant required to run for grid reliability
ROR	Rate of Return, calculated on total assets or on investors' equity
SC	Scheduling Coordinator, in California an energy market that submits balanced schedules to the system operator
SNT	Short Notice Transaction, in New England a supplier's action to withdraw its bid from the real-time stack of incs and decs; e.g., to export to NY
SO	System Operator, often denoted ISO to convey functional independence
TCC	Transmission Congestion Contract, a financial hedge against transmission usage charges
T&D	Transmission and Distribution, referring to the "wires" businesses
TO	Transmission Owner
UDC	Utility Distribution Company, a regulated retail distributor of energy over its local lines, a synonym for LDC.
VicPool	The system operator in Victoria before formation of the ANG
WEPEX	Western Power Exchange, an energy market operating in the western region of the U.S. and Canada prior to restructuring
WSCC	Western States Coordinating Council, an organization that establishes regional reliability standards and procedures

Appendix B: Dictionary of Technical Terms

Adjustment market – The market for Incs and Decs used to adjust generation to ensure that transmission constraints are not exceeded.

Ancillary services (AS) – Capacity reserved to ensure transmission reliability and for real-time load following. Typical categories are automatic generation control (AGC) for load following on a second-by-second basis, spinning reserve capacity that can be ramped in 10 or 30 minutes, and replacement reserve capacity that can be called with 1 hour notice to replace spinning reserve that has been activated. Reserves are sometimes differentiated between incremental and decremental reserves. Common acronyms are AGC, TMSR (spin that can ramp in 10 minutes), TNSR (non-spinning reserve that can ramp in 30 minutes or 1 hour), ? (operating or replacement reserve).

Auction – In an exchange, a market conducted by receiving bids from suppliers and demanders and then using a market clearing procedure to establish the price that equates supply and demand. Sometimes termed a "call" auction, or a "double" auction when referring to bids from both suppliers and demanders. Some auctions, called "bid-ask markets" and sometimes "double auctions," provide a forum such as an electronic bulletin board for posting of bids and offers that can be accepted or negotiated bilaterally at prices that differ among transactions.

Balanced schedule – A schedule of energy generation in which injections into the transmission system equal the sum of withdrawals and losses.

Basis – The price difference between a specified location and the reference hub. The price at the location is the sum of the price at the hub and the basis. Basis risk pertains to uncertainty about the magnitude of this difference.

Bilateral market – A market in which trades are contracted directly between a seller and a buyer at negotiated prices, terms, and conditions specified in a contract between the two parties. Some bilateral markets are mediated or conducted by brokers and dealers; others provide forums for posting bids and offers, often on an electronic bulletin board.

Black-start reserves – Generation capacity capable of restarting after a system collapse without requiring any external source of energy.

Blue Book – The 1994 Notice by the California PUC that initiated restructuring of the electricity industry in the U.S.

Broker – An intermediary who facilities trades between other parties without taking a financial or physical position, in contrast to a dealer who does.

Bus – A point of receipt into or delivery from the transmission system, essentially synonymous with node in the parlance of electricity markets.

Cascade markets – Markets for ancillary services in which bids rejected for a higher quality reserve, such as 10-minute spin, are re-entered as bids for the next lower quality, 30-minute spin.

Common carriage – The body of principles derived from statutes, precedent, and regulatory policies regarding the transportation industries that set standards for open access transmission on nondiscriminatory terms.

Competitive transition charge – The non-bypassable retail surcharge on energy used in the period 1998-2002 to pay off debt obligations of utilities for prior investments.

Compliance – Adherence to the procedural rules of a market or the system operator. Noncompliance is detected by testing or auditing, and sanctioned according to rules and penalties in the tariff.

Congestion – The event that initially scheduled flows on transmission lines exceed safe limits required for reliability, stability, or protection against risks of line or generation failures elsewhere in the system.

Congestion management – Incremental and decremental adjustment of generation or loads in various locations or zones to ensure that transmission capacity constraints are not exceeded.

Contestability – Mitigation of market power by easy entry of competitors. Examples are imports enabled by large transmission capacity, and investments in new generation equipment by entrants.

Contract for differences (CFD) – A bilateral contract with a specified strike price in which each party ensures the other against discrepancies between the strike price and spot prices.

Counter-party – The party with whom a contract is concluded, such as the exchange when it acts as intermediary for all transactions. In bilateral markets, counter-party risk refers to the chances that the counter-party defaults on its obligation.

Dealer – A market intermediary who buys and sells for its own account, usually based on taking risky positions. As used here, a power exchange that acts as a counter-party to each transaction is not a dealer because it does not take a position; that is, equality of demand and supply at the clearing price results in no net position for the exchange, and no profits from trading.

DC approximation – Representation of an alternating current system by an approximation based on direct current. The DC approximation has simpler mathematical properties than the true alternating current formulation.

Default usage charge – The transmission usage charge imposed by fiat, subject to FERC approval, when there are insufficient adjustment bids to alleviate transmission congestion.

Dispatch – Instruction from the system operator regarding the operation of a generator.

Entitlement – A contractual claim to a share of the energy output of a generator. Jurisdictions such as Alberta auction entitlements to the output of large incumbents to mitigate their market power.

Exchange (PX) – As used here, an energy market in which supply offers and demand bids for each standard traded commodity are aggregated to find a clearing price at which total supply and demand are equal. This usage here is adopted to distinguish exchanges from pools with optimized dispatch and broker or dealer markets for negotiated bilateral contracts. Also connotes a market for financial instruments, such as a futures exchange like NYMEX or commodities exchange like MERC.

Firm transmission right (FTR) – In California an FTR comprises a transmission congestion contract (TCC) plus assurance of scheduling priority on a *pro rata* basis if there are inadequate adjustment bids to alleviate congestion.

Flexible resource – A generation source, such as a combustion turbine, that has a low cost of startup and a high ramping rate.

Futures market – A market, such as NYMEX, for financial instruments conditioned on delivery at a particular time and place. For example, one NYMEX contract specifies delivery at the California-Oregon Border (COB) and another at Palos Verde (PV). Futures contracts are often cashed out without actual delivery.

Gaming – Behavior by a participant intended to exploit imperfections in the design of a market, especially loopholes in the procedural rules.

Governance – The system established for directing the affairs and policies of a market or system operator. Includes a Board of Directors, or Governors, often with wide representation from stakeholders, that appoints and supervises management and that has ultimate responsibility to state and federal regulatory agencies.

Hedge – A financial instrument that insures the buyer the events against subsequent events, such as price variations. Examples are a futures contract or a swap. Some jurisdictions, such Alberta and England-Wales in their initial years, rely on legislated hedges to mitigate market power by altering the incentives of large incumbents.

Hub – A location in a transmission grid used as a reference point for pricing.

Intervenor – A party that files comments or protests in regulatory proceedings.

ISO – A system operator with substantial independence from market participants. Later usage by FERC relies on the term RTO for regional transmission operator.

Incentives – A participant's perception of rewards for adhering to the procedures and rules of a market. A market design is called incentive-compatible if participants are motivated to adhere to the specified rules and procedures.

Incs and Decs – Bids from suppliers and demanders to increase or decrease generation or load.

Kirchhoff's Laws – The classic theoretical statement of the physical laws governing flows in an electrical transmission system. Basically, the flow of electrons among multiple paths from an origin to a destination splits in inverse proportion to the impedance of each path.

Locational marginal pricing (LMP) – Pricing energy at each node in the transmission system, also called nodal pricing. Prices differ at locations when transmission between them is congested.

Loop flow – Flows in one part of a transmission system that originate elsewhere but follow paths dictated by Kirchhoff's Laws.

Losses – Energy expended in transmission, lost as heat. Sometimes approximated by a generation meter multiplier (GMM).

Marketer – An entity that buys energy wholesale for resale to retail customers.

Market clearing – The process in an exchange of finding the price that equates supply and demand. The resulting transactions comprise the supply offered at lower prices and demand bid at higher prices.

Market clearing price (MCP) – The price that equates supply and demand.

Market failure – Absence of a price at which a market clears. An example is the adjustment market when the incs and decs offered are insufficient to alleviate transmission congestion. Another is a market for reserves when there are insufficient bids to supply the reserves required by engineering standards for reliability. Sometimes called bid insufficiency.

Market power surveillance – Measures undertaken by exchanges and system operators to monitor the behavior of participants with substantial market power, and to recommend measures to mitigate or control market power. Required by FERC in some jurisdictions. In California the responsibility is divided among Market Surveillance Units (MSUs) in the PX and the ISO, as well as independent Market Surveillance Committees (MSCs) that rely in part on the MSUs to provide administrative and technical support services.

Merit order – The ordering of generation units in a sequence of increasing marginal costs. Absent transmission and ramping constraints, units are called to generate in merit order.

Market surveillance committee – A committee that in California is charged with monitoring the ISO's markets for abuses of market power and other sources of inefficiency. In New Zealand the MSC is a judicial body charged with enforcing the pool's enabling contract among the participants.

Must-take supplies – Energy produced by plants that incur large costs of shutdown and startup, such as nuclear facilities, is offered in an exchange at a zero price to ensure that it is accepted in the market clearing. Must-serve demands are analogous.

Nodal price – In systems that set prices for energy at each location, the price of energy at a particular location (a node) or bus.

No-load cost – The hourly fixed cost of operating a generator, apart from startup cost and the marginal cost of energy generation.

Nomination – In gas markets, the term for a schedule submission, which in the case of interruptible capacity is accompanied by an offered price, so the nomination is essentially a bid as the term is used in electricity markets. Gas-fired generators often depend on nominations for gas delivery, and some jurisdictions require the system operator to call generators sufficiently early to enable a nomination to be submitted in the day-ahead market for gas transmission. Markets for other commodities use a variety of terms for bids.

Optimal power flow program (OPF) – A computational procedure for spatial allocation of generation to meet load requirements at least cost subject to transmission constraints.

Option – A financial contract enabling the holder to buy a commodity at a specified price, called the exercise price, or a physical contract enabling the holder to control the operation of a plant, as in the case of RMR contracts that enable the system operator to call for energy production from a plant in order to ensure system reliability.

Order – The decisions and rules promulgated by FERC in a formal statement binding on those entities subject to its jurisdiction. Examples are Orders 888 and 889 that implement the 1992 Energy Policy Act and set the main rules for restructured electricity markets based on open access to transmission and non-discriminatory pricing.

Pool – Traditionally, a consortium of utilities conducting unified generation and transmission operations. As used here, a system operator that optimizes unit commitments and schedules based on submitted offers from suppliers and possibly also bids from demanders.

Protocol – The specific procedures used by a system operator or power exchange, usually elaborated in detail in the tariff.

Pumped storage – A hydro source whose supply is filled by pumping water from lower elevations, usually in off-peak periods.

Radial system – A transmission system that is largely immune to loop flows because it consists mainly of spokes from a hub. Jurisdictions such as California use a radial approximation for day-ahead congestion management.

Ramping up or down – Increasing or decreasing the production rate of a generation unit. Ramping stresses the turbine and boiler of a thermal generator. The maximum rate at a which a generator can be safely ramped is called its ramp rate. Typical ramp rates are on the order of 1% of maximum capacity per minute. Some hydro units are nearly instantaneous but are subject to a total energy constraint, referring to the maximum feasible spillage per day.

Rational buyer procedure – Selection of the amounts of various ancillary services reserve categories to minimize total cost, using the fact that a higher quality reserve can substitute for lower qualities. For example, spinning reserve that can ramp in 10 minutes is superior to spinning reserve that can ramp in 30 minutes, so if 10-minute spin can be purchased at a lower price than 30-minute spin then it is cost minimizing to substitute the former for the latter.

Reactive energy – ? (VAR)

Real-time market – The continual market for energy to follow the load and to maintain reliability of the transmission system.

Regulation – Either (a) supervision by a state or federal agency empowered to implement statutes governing the industry, or (b) the ancillary service called automatic generation control (AGC) that enables near-instantaneous load following by a generator.

Regulatory compact – An implicit understanding between regulatory agencies and regulated utilities about the *quid pro quo* in which the utilities are assured a monopoly and an adequate rate of return on assets or equity for prudent investments, in exchange for fulfilling service obligations at regulated prices and terms.

Reliability-must-run plants (RMR) – Plants located at critical nodes in the transmission grid that sometimes must operate to ensure stability of the system.

RMR contract – One of several long-term contracts approved by FERC under which the system operator can call for generation from plants required for system reliability.

Renewable resource – A source of generation that does not use fossil or nuclear fuels, such as wind power, geothermal steam, heat from biomass or waste burning, or small hydro facilities. Major hydro sources are often excluded.

Retail competition – Restructuring of retail markets for power that enables each customer to choose its energy supplier, possibly a power marketer or a utility distribution company.

Run-of-river – Hydro power from spills required to accommodate large flows of water into a dam, usually in the spring thaw.

Scheduling coordinator (SC) – In California, an entity entitled to submit balanced schedules to the system operator. The main ones are the PX and APX exchanges, managers of markets for bilateral contracts, and some municipal utilities.

Security analysis – The system operator's process of checking that the transmission system can survive the failure of major links in the grid. If the survival criterion is restricted to failure of any one line then the process is called an n-1 analysis. The process may include determination of loop flows and examination of additional contingencies such as temperature or wind variations. Because security analysis takes considerable time, it is often omitted from the initial day-ahead management of congestion and deferred to the interim before real-time operations.

Self-management of congestion – Measures undertaken by a market participant at its own expense to reduce transmission congestion, usually with the intent of reducing the usage charge for transmission.

Self-provision – Supply of various ancillary services by a supplier or demander rather than the system operator.

Self-scheduling – Unit commitment and scheduling decided and controlled by the unit's owner rather than the system operator.

Settlement – The process after the close of a market in which payments are made at the market prices to complete the transactions or contracts. Sometimes interpreted to include *ex post* metering of actual quantities sent or delivered.

Shadow price – In the results obtained from minimizing total cost subject to feasibility constraints, the shadow price on a constraint is the decrease in total cost enabled by a small relaxation of the constraint. If the constraint requires that supply at a particular time and location must exceed a specified demand measured in MWh then the shadow price measured in $/MWh is the reduction in total cost if demand is decreased by 1 MWh. Similarly, if the constraint requires that transfers not exceed the capacity of a transmission line in a given hour then the shadow price measured in $/MWh is the reduction in total cost if the capacity is increased by 1 MW for that hour. Shadow prices are called dual variables in the theory of linear programming and optimization generally. They measure the opportunity cost of serving one customer versus the next in the merit order.

Short-notice transaction (SNT) – In New England a transaction submitted shortly before dispatch that alters a supplier's generation within the system, such as export to another control area like New York or shutdown of a generator, that has the effect of removing all or a portion of the supplier's capacity from the merit order for real-time dispatch. Other jurisdictions such as Alberta allow a short-notice alteration of a supplier's schedule, called a redeclaration.

Synchronization – The process of connecting a generator to AC transmission system to assure compatibility of the frequency and phase angle.

Spin – Capacity reserved to meet contingencies, usually classified according to the ramping time required, such as 10 minutes or 30 minutes.

Spot price – In an exchange the market clearing price in the real-time market, or in an optimized system the shadow price on the constraint that supply equals demand in real-time operations.

Stakeholder – A party or organization with a substantial interest in the design and operation of a market or system operator, often formally recognized via participation in design debates and on governing boards.

Stranded cost – The difference, if positive, between the book value and the market value of a plant divested by a regulated utility as part of the restructuring of a wholesale market.

Strip – A contract for energy delivery, usually at a constant rate, over a span of consecutive hours, often defined as the peak or off-peak hours. Strips are traded in some brokered markets, often in the form of weekly contracts for 5 daily strips covering the peak hours of the weekdays.

Swap – A trade of energy at one location for energy at another location, often used as a hedge against transmission charges.

System operator (SO) – The entity responsible for the reliability and security of the transmission system within a specific control area is called the system operator. The system operator is a public-benefit corporation in California, a creation of the local power pool in New England, and a regional utility or government agency (BPA) in Oregon. Municipal utilities often manage their own grids within the control area of the system operator.

Tariff – In a restructured electricity market each regulated entity, such as the system operator or an exchange, files with FERC a tariff that specifies in detail its governance and operating procedures, terms and conditions, charges, and pricing policies.

Thermal rate – A thermal generation unit's rate of conversion from fuel energy to electrical energy.

Tie line – A major transmission artery.

Transaction charge – The fee charged by an exchange, broker, or system operator for completing a trade in its market.

Transmission congestion contract (TCC) – A point-to-point TCC is a financial contract that pays the owner the amount of the usage charge for transmission on that path. Used by suppliers and demanders to hedge against transmission usage charges or differences in nodal prices.

Unit commitment – Designation of which generation units will operate, usually for the ensuing day, often interpreted to include times of startup and shutdown as well as operating rates.

Uplift – A surcharge on transactions that repays the system operator for those costs incurred that are to be shared *pro rata* by traders. Examples are costs of alleviating congestion (when congestion charges are not used) and procuring ancillary services.

Usage charge – The fee charged by the system operator or the owner of the transmission line for an energy transfer. May include an access fee and/or a flat rate, but includes a charge for the costs of alleviating congestion only if some form of congestion pricing or nodal pricing is used.

Zone – A region in which no congestion charge is imposed for transmission; usage charges are confined to inter-zonal transmission.

Appendix C: Effects of Market Structure on Reliability Management

The structure of the markets for energy can have substantial effects on the measures used by the system operator (SO) to ensure reliability of the transmission grid. This section describes some of the interactions between reliability management and energy markets. We start with examples that illustrate the need for caution when introducing seemingly convenient features into energy markets that complicate the task of the SO – and vice versa.

Coarse Time Scale. California is one of several jurisdictions that conduct day-ahead and hour-ahead markets for energy on the basis of each of the 24 hours of the day of delivery. Day-ahead transactions in the power exchange (PX) are settled on the basis of the market clearing prices (MCPs) that equate demand and supply in each hour of the next day, measured in terms of total MWhs. The PX operates on the principle of self-scheduling, so each supplier makes its own decisions about unit commitments and unit ramping to meet the commitments implied by its sales of energy. A further aspect is that uninstructed deviations from prior schedules contracted day-ahead or hour-ahead are settled at the real-time price. This scheme creates a difficulty for the SO in shoulder hours when the load varies greatly, increasing rapidly in the morning and decreasing rapidly in the evening, on a time scale considerably shorter than the hourly time scale of the PX's energy markets. One can imagine the PX transactions as based on a step function of time that in each hour specifies a constant rate (MW) of energy supply, whereas in real-time operations the intra-hour time path of supply must vary to match the load. Further, for suppliers there is an evident need, say in the morning, to ramp through the first shoulder hour to meet its supply commitment in the next hour. Thus, the SO lacks tight control in the shoulder hours of the time-path of supply required to meet the load, and at any moment the actual supply from committed generators can be greater or less than the load. The SO's real-time market for load following and intra-zonal transmission management is based on supplemental energy bids (incs or decs) that might be used to cover the discrepancy between the load and generation, but in fact if there is any doubt about the sufficiency of the bids available in the real-time market then the SO is obligated to rely on some kind of reserve. Absent a new category of reserve for load following in shoulder hours, the SO in California often relied on its reserve of automatic generation control (AGC) for this purpose. Instead of purchasing the 2 or 3% of AGC capacity common in earlier eras, the SO purchased as much as 12% AGC for the shoulder hours in order to ensure adequate capacity for load following. This extra reserve in shoulder hours can be interpreted as a consequence of the hourly market for energy in the PX instead of the 30 or 15-minute time frame that would conform closer to the time scale in which load varies in shoulder hours. The fact that AGC is used in California for this purpose, rather than another (or new) category of reserves, is of less significance that the fact that the origin of the problem lies in the "coarse" time scale of the basic energy markets compared to the time scale of load variation in shoulder hours. One proposal to reduce this problem is to redefine the contracts traded in energy markets such as the PX so that, rather than a step function of time, the energy commitments are interpreted as piecewise-linear functions of time that include a specified rate of ramping from one hour to the next.

Asymmetric Treatment of Demand and Supply. The day-ahead operating procedures of tight power pools were designed to minimize the cost of supplying the predicted load. An SO with operating protocols derived from an earlier era may prefer to continue this practice even after demand-side bidding is introduced. A typical scheme proposes to schedule supply to meet the predicted load even if day-ahead demand bids are less than the SO predicts will materialize in real time. Thus, the bid-in day-ahead demand is contracted at the price that obtains enough supply to meet the predicted demand, and the residual real-time demand is settled at the real-time price. At first sight this scheme protects the SO against reliability problems from day-ahead under-bidding of demand by ensuring a supply adequate to meet the subsequent total demand.[62] In fact, however, it vitiates the potential efficiency advantages from demand-side bidding. To see this one must realize that for a demander this scheme implies that its optimal strategy is to bid in all its demand at its prediction of the real-time price: if the day-ahead price is lower then its bid is accepted at that lower price, and if the day-ahead price is higher then its bid is rejected and it pays the real-time price, which is likely lower in its estimation. Thus, the bids from demanders reflect only their optimal strategies of gambling on discrepancies between the day-ahead and expected real-time prices; there is no systematic revelation of the quantities ultimately demanded at each price. The source of this problem is that the various demanders' estimates of the quantities and prices in real time typically differ from the SO's predictions. It seems fundamental that such problems arise whenever the SO attempts to ease its reliability problems by "out-guessing" demanders, rather than relying on actual bids and equalization of day-ahead and expected real-time prices by normal arbitrage. It is the responsibility of the SO to schedule additional reserves if it suspects underbidding could undermine reliability, but this solution is better than attempting direct scheduling of supplies to meet predicted demands in ways that affect the market clearing prices for energy in the day-ahead market.

Unbalanced Day-Head Schedules. Scheduling supplies to meet the predicted load, as above, is one example of day-ahead schedules that are unbalanced; that is, the supply contracted day-ahead exceeds the demand contracted day-ahead. Another example is scheduling of reserves to meet local reliability needs, as in the case of "reliability must run" (RMR) units in California. A fundamental deficiency of the market design in several jurisdictions is that they establish prices for energy and transmission that equate demand and supply aggregated over the entire system or large zones, when in fact the SO requires adequate generation at certain key points in the grid to maintain reliability. Recognizing that local requirements convey market power to suppliers in these locations, the California SO obligates suppliers with localized market power to operate under long-term contracts that provide cost-based payments in exchange for the SO's right to call for

[62] Anxiety about underbidding by demanders might stem from concern that large LDCs with monopsony power might underbid day-ahead to lower the prices paid for that portion of their total demand. This concern must, however, be balanced against the prediction that the day-ahead price and expectation of the subsequent real-time price must be equal, because otherwise suppliers would prefer to withhold some supply in order to obtain a higher expected price in the real-time market.

generation when needed. In 1998, about a third of the intra-state thermal capacity was assigned to RMR contracts. The contract forms available at that time were imperfect in part due to distorted incentives, but here we focus on the consequences for reliability.[63]

The central problem for the SO stemmed from the fact that the RMR units were called after the close of the day-ahead energy markets. This was thought to be the better procedure because then the units called could be confined to those not already committed due to their transactions in the day-ahead market; in fact, however, the contractual incentives encouraged most of the RMR units to wait for the call from the SO. Because each of the energy markets, including the PX, submits a balanced schedule to the SO, the additional energy from the RMR units produced a supply surplus in the day-ahead schedule, a surplus of energy that was not matched by any load, implying elevated day-ahead prices compared to the actual energy supplied. When the real-time market arrived, therefore, it was possible that the system was awash in energy, forcing the SO to decrement suppliers to balance the system, and thereby producing low real-time prices. The evidence, however, is that suppliers and demanders reacted to this potential price discrepancy by arbitraging between the day-ahead and expected real-time prices. In particular, demanders deferred a large portion of their purchases to the real-time market. The ultimate consequence for the SO was an aggravated reliability problem in the real-time market. Instead of a small residual market for intra-zonal balancing and load-following, the SO's real-time market was a major market for trading energy.

The solution to this problem was to change the contract specifications. The most important changes were for the SO to call RMR units in advance of the day-ahead markets, based on its forecasts, complemented by incentives for the RMR suppliers to sell their called energy through the day-ahead markets. Thus, the balanced day-ahead schedules submitted by the energy markets to the SO included most of the RMR energy.

The second and third of these examples convey a lesson about limitations on the measures available to the SO to maintain the reliability of the grid. The lesson is that measures adopted in one market, say the day-ahead market, cannot ignore the incentives of suppliers and demanders to take advantage of any systematic price discrepancies between that market and later markets, including transmission, ancillary services, and the real-time market. Basically, one must suppose that arbitrage will ensure approximate equality between the day-ahead price and the expectation of the subsequent real-time price, because traders can defer sales or purchases. The most the SO can do is to affect the relative magnitudes of the quantities traded in the two markets. For reliability, the SO prefers the bulk of energy trades to be transacted day-ahead so that the imbalances arising in the real-time market are sufficiently small to be dealt with easily and inexpensively. Measures to force this outcome cannot work if they rely on sustained discrepancies between prices in the day-ahead and real-time markets.

---

[63] See Frank A. Wolak and James Bushnell, "Reliability Must-Run Contracts for the California Electricity Market," Market Surveillance Committee, California ISO, 10 December 1998. [Cite also the PricewaterhouseCoopers audit of the ISO, Nov. or Dec., 1998; available on the ISO website.]

Appendix D: Comparison with Gas and Pipeline Markets

The structural features of broadest significance in restructured electricity markets are one or more power exchanges (PXs) for trading energy and a system operator (SO) for managing the transmission system. These are separate in jurisdictions such as California and unified in others such as New England (NE) and Pennsylvania-New Jersey-Maryland (PJM). The designs in the U.S. are derived from those implemented in other countries, such as Australia and Scandinavia in the case of California, and England-Wales in those cases with unified operations. Congestion pricing of interzonal transmission derives from Scandinavia, but PJM's pricing at the nodal level is new – although it is based on well established theoretical models. As in most other countries, the PX and the SO are public enterprises or public-benefit corporations governed by a Board that includes representation from stakeholders. Arguments have been made that transmission management might be a franchise operated by a private corporation with explicit incentives designed to enhance efficiency, but no jurisdiction has adopted this approach.

These structural features have been adopted in some other countries with restructured markets for natural gas and gas transmission via pipelines. The designs in Victoria (Australia) and the U.K. are similar to those used for electricity markets, excepting only those aspects peculiar to gas such as the availability of storage, the ability to direct flows point-to-point, and a different time frame (monthly/daily rather than daily/hourly). In the U.S., however, the markets for commodity (gas) and transmission (pipelines) are based on the principles of "contract carriage" that were designed originally to promote construction of new pipelines by private corporations.[64] The regulation of pipelines is different, moreover, since it consists mainly of a maximum price that a pipeline is allowed to charge.

In this section we describe the key features of gas and pipeline markets in the U.S., and compare these markets with the restructured electricity markets for energy and transmission. Our main conclusions point to the evident inefficiencies of pipeline markets and their origin in the pervasive market power of the pipeline companies.

What's Different About Gas and Pipelines

Unlike electricity flows that are governed by Kirchhoff's Laws, gas flows can be directed point-to-point using valves. Rather than traveling at the speed of light, gas flows at about 50 mph, although this rate is sensitive to pressure and temperature. The pipeline owns the linepack (gas stored in the pipeline), which can be varied by altering the pressure, but customarily a point-to-point transmission is considered to occur instantly; that is, transmission is interpreted as "displacement". Interpreting the linepack as incompressible, an injection at a receipt point and a balanced withdrawal at a delivery point are considered to be simultaneous. Unlike electrical SOs, pipelines do not provide credit for counter-flows, and indeed charge for them, and also charge for both parking (temporary storage in the linepack) and lending (temporary withdrawals from the

---

[64] Although electrical transmission is not based explicitly on principles of common carriage, the effect of FERC Orders 888 and 889 is essentially to establish analogous principles.

linepack) even though it is one customer's parked gas that is loaned to others – that is, there is no spot market for balancing. Even though gas is uniform, the industry has an elaborate system to trace exactly the ownership of gas "from wellhead to burnertip."

The markets for gas are entirely private and entirely bilateral; there are no central exchanges other than the NYMEX market for futures contracts. There are many brokers, and some producers have marketing subsidiaries, but many producers sell only to pooling points in supply areas. Some marketers can choose among two or more pipelines to major market areas, and in any case one can choose among swaps offered by brokers. A typical swap trades gas at one point for gas at another point. Transmission and storage accounts for as much as half the delivered price of gas, as compared to a small fraction for electricity. The variable cost of transmission is significant, due in part to the operating costs of compressors spaced along pipelines.

A large fraction of demand is from local distribution companies (LDCs), and because most is for winter heating, demand is strongly seasonal, peaking in the months November to March. The peak load is served mainly from underground storage near "city gates" in market areas. Consequently, LDCs tend to contract long-term for both gas supplies and transmission reservations, and a major portion of transmission is used year-round to fill storage vaults in market areas; storage in supply areas is relatively scarce and short-term. Some industrial loads, especially gas-fired electrical generators, have pronounced intra-day peaks.

In contrast, the supply from wellheads is essentially constant except for occasional events such as deep freezes. Moreover, the supply from each wellhead is constant, and producers consider maintenance of constant flow to be of paramount importance. This is because interruption of the flow defers an entire stream of revenues from the gas, and often more important, the oil behind the gas; and more important still, shut-in can permanently injure the long-run productivity of the well.

The basic planning cycles in the gas market include multi-year and seasonal contracting for supplies, a major market in the "bid week" preceding the beginning of each month, and a daily spot market. Many of the short-term transactions are pegged to the NYMEX futures contracts for delivery at a few major hubs, or to spot prices at these hubs, even if the gas is not routed through these hubs.[65]

The demand for gas is expected to grow significantly, perhaps a third by 2010, with a corresponding need for new wells and expansion of pipelines, even though overbuilding of pipelines is perceived as a significant factor presently on several routes.

The Transmission Markets

The most distinctive aspect of gas markets is the central role of the separate markets for transmission conducted by the pipelines. The immediate cause is the dependency felt by

[65] Because the profitability of a marketing operation is typically measured against an index of hub spot prices, a long-term fixed-price contract is often seen as risky, in contrast to the usual view that locking in a price hedges against price risks.

captive suppliers who fear the high costs of shut-in, and the LDCs' service obligations, although as deregulation of retail gas markets proceeds at the state level LDCs are likely to be less involved in purchasing, concentrating on distribution as in California where deregulation has proceeded rapidly. The long-term cause, however, lies in the regime of contract carriage developed to encourage pipeline construction.

Producers' fear of shut-in and LDCs' service obligations coincide with the pipelines' need for capital for construction or expansion. These mutual interests are served by the long-term contract known at FT, meaning firm transmission. FT obligates the shipper to pay a charge, called a demand charge or a reservation fee, each day for the reserved capacity whether it is used or not, plus a volumetric usage charge for actual flow that is close to variable cost. The pipeline uses these obligations as security to obtain debt and equity financing for construction. Regulation consists mainly of setting a maximum price (for each specified pair of receipt/delivery, expressed as $/mcf) at which FT can be sold, and an obligation to sell at this price if capacity is available (a pipeline can accept a lower price if it wants). At the expiration of an FT contract the shipper has a right of renewal at a term up to 5 years, but for pipelines the major concern currently is "turnbacks" (i.e., non-renewals) by LDCs anticipating state deregulation that would remove their service obligations. FT is not financially firm for shippers, since the demand and usage charges can be increased in a rate case filed with FERC by the pipeline, but in practice most pipelines prefer not to file rate cases (and others challenging rates must bear the burden of proof when in fact they have little information about the internal affairs of pipelines) so that they can exceed their allowed rates of return in years between rate cases.[66]

There are two secondary markets for transmission. One is the secondary market for FT. Although it is an active market, mainly on a monthly or intra-month basis via an electronic bulletin board (EBB) maintained by the pipeline pursuant to FERC directives, resale is encumbered in three ways. One is that resale is subject to the same price cap. A second is that the FT contract specifies "primary points" for receipt and delivery, whereas a buyer often wants other "secondary points" that are available only at the discretion of the pipeline. The third is that the seller, usually an LDC, imposes recall provisions that substantially impair the firmness of the transmission reservation. Indeed, the intra-day market consists mainly of reallocation of capacity in response to LDCs' recalls, leaving the secondary purchasers to find other capacity available on an interruptible basis.

The other secondary market is for interruptible transmission (IT), sold each day on a space-available basis. A peculiarity of the U.S. market is that a pipeline typically has some unused capacity available for sale as IT that was previously sold as FT but has not been scheduled; this anomaly occurs because regulated LDCs can pass on the cost of FT to their ratepayers and so some have little incentive to participate actively in the secondary market for resale. However, the pipeline currently has full discretion to decide how much IT to sell, even if its capacity is not fully booked via FT contracts.[67] IT is sold

[66] Some studies indicate rates of return for a few pipelines nearly twice the allowed rates, and sales of pipelines have occasionally been at prices up to double the replacement cost.

[67] This situation, in which pipelines withhold available capacity to obtain higher prices for IT, is the motive for FERC's Notice of Proposed Rulemaking, July 1998, that suggests requiring each pipeline to offer all

daily via bids, called nominations, submitted by shippers: the pipeline decides which point-to-point nominations to accept, and winning bidders pay their bids. The pipelines' allocation procedures are opaque but presumably those point-to-point nominations are accepted that maximize the pipeline's revenue subject to operating constraints and the amount of IT capacity that the pipeline chooses to make available.

Taken together, FT and IT contracts represent a system of priority pricing. FT has first priority in scheduling, subject only to *pro rata* rationing in the event that it is infeasible to serve the entire demand from FT contracts (some pipelines sell more than 100% of their certified capacity as FT, with the portion above 100% having secondary priority). The financial terms for FT differ among pipelines: the rates on two pipelines serving nearby corridors can differ substantially due to differences in their maximum rates (due to differences in embedded costs), differences between point-to-point and zonal pricing, and procedures (e.g., some pipelines use auctions). The lowest priority is IT, which is available only in amounts chosen by the pipeline, and only on a daily basis after all FT has been scheduled. Other than these differences between FT and IT, the remainder of the terms and conditions are specified in the pipeline's tariff filed with FERC.

A pipeline whose certified capacity is fully subscribed by FT contracts bears essentially no financial risks because demand charges are payable in any event, and can be increased if justified in a rate case, even in cases of turnbacks in which remaining shippers are charged more to obtain revenues previously paid by departing shippers. A pipeline also has additional profit opportunities from cost cutting, sales of auxiliary services such as parking and lending, and sales of IT, which account for rates of return above the rates allowed in the previous rate case.

For a shipper, a long-term FT contract is a bundle comprising a physical contract for daily scheduling priority and a financial contract that refunds the daily IT spot price (actually the excess of the minimum accepted bid over the FT usage charge) necessary to obtain transmission. It also includes a right of resale (with recall provisions enforced by the pipeline) and at expiration a right of renewal. The price of this contract is allocated over the term of the FT contract via the daily payments of demand charges. The main financial risks are the chances that IT will turn out to be cheaper, or that a rate case will increase demand charges (although the pipeline likes to mitigate both these). These risks are offset by eliminating the risk that transmission will be unavailable at any price when the pipeline offers no daily IT at all. This is offset is especially important to producers worried about shut-in and LDCs facing a service obligation at fixed prices.

The Effects of Monopoly Power

Other countries such as Australia and the U.K. have eliminated the monopoly power of pipelines by establishing a system operator with an obligation to serve the public interest. The U.S. rejected this approach in previous eras and recently FERC's initiatives to address its faults have reiterated an intention to strengthen further the role of long-term

---

available capacity as IT in a daily auction without a reserve price, but also without a price cap. At the same time FERC issued a Notice of Inquiry in which it proposed to strengthen the role of long-term contracting for FT.

contracting for FT, presumably to ensure sufficient capacity expansion and to protect the financial security of existing pipelines facing turnbacks. FERC may also mandate daily auctions of IT with no price cap and no reserve price for all available capacity not scheduled by owners of FT. Such an auction is intended to eliminate the inefficiency evident in the pipelines' withholding of available capacity in order to maintain prices for IT, and thereby also FT.[68] The source of this inefficiency is each pipeline's market power.

A pipeline's market power stems from several sources. One, of course, is the scarcity of physical substitutes. Unlike liquids such as oil, alternative modes of transport are infeasible, and many supply and market areas are served by only one or a few pipelines. Entry of new pipelines is so costly and time-consuming, not to mention the difficulties of acquiring rights of way, that there is little prospect that the market is contestable. The terms of financial substitutes such as futures and swaps are dictated largely by the price of transmission. A second source is the pipeline's ability to withhold FT capacity at any price below the maximum price, and equally to withhold capacity from the daily market for IT. A third source is the bundling in FT of physical scheduling priority with a financial hedge against the prices of daily IT: a shipper cannot obtain the highest level of scheduling priority unless it subscribes to FT. Also, a shipper cannot buy a financial hedge: no potential seller of a competing hedge could plausibly insure against IT prices that are determined by the pipeline's daily choices about how much IT to sell, and where the purchaser pays its bid price offered in ignorance of the minimum price required for its nomination to be accepted – indeed the lack of transparency in the procedures for selling IT, and the pipeline's private knowledge of how much capacity is potentially available, render infeasible any external market for financial hedges against IT prices. A pipeline obtains further profits from its monopoly on parking and lending and other auxiliary services. A fourth source is the pipeline's ability to defer a rate case until a time of its choosing.

An important question is whether FERC's proposed daily IT auction of all available capacity can be made to work, and whether in fact it would mitigate the market power of pipelines. At the very least it would require new measures to bring standards and transparency to the process, to force the revelation of how much capacity is actually available, and to prevent bypass in which the pipeline avoids an auction by selling all its capacity to affiliates or third parties who can control resale or encumber resold FT with recall provisions and inferior primary points. This is not easy when capacity is a many-faceted concept and nominations are point-to-point. But rather than dwell on this question here, we turn to some lessons for restructuring wholesale markets for electricity and transmission.

Implications for Restructuring Electricity Markets

Much controversy about electricity restructuring focuses on operating procedures of the energy markets and the SO's procedures for managing the transmission grid. It is easy in such debates to ignore the dominant fact that perhaps the largest efficiency gains result

---

[68] A further effect is elimination of favoritism towards affiliates, which is a major problem in this industry, evidenced most clearly in the sanctions imposed on the Natural Gas Pipeline in 1997-8.

from the exclusion of the monopoly power that would otherwise accrue to private owners of transmission lines and market institutions. The most glaring inefficiencies in gas transmission – capacity withholding, charges for compensating parking and lending from the linepack, exclusion of a balancing market, charges for counter-flows, high charges for interconnections with other pipelines, lack of coordination among pipelines, etc. – would presumably occur also in electrical transmission if a similar regime were adopted. These inefficiencies and the persistent problems (e.g., favoritism to affiliates) stemming from the market power of pipelines – even though nominally regulated by the same federal agency – attest to the fundamental advantages obtained from assigning transmission management to a system operator with an obligation to serve the public interest, or at least the interests of stakeholders represented in its governance structure.

However, a design that relies on a benign system operator could bring other difficulties in the future. The source of the most critical problem, capacity expansion, lies in the governance structure. A new line typically raises export prices at the receipt end and lowers import prices at the delivery end. At each end, the interests of suppliers and demanders are opposed. For example, in provinces like Ontario and Alberta where one or a few firms own most of the generation, an expansion of import capacity would enable external suppliers to erode the market power of incumbents; or in states like Oregon with low-cost hydro sources, an expansion of export capacity to California would tend to equalize consumers' prices in the two states. When broad support from stakeholders is required for approval of capacity expansion, the likelihood that these opposing interests are reconciled may be small. The problem is further complicated, moreover, by the lack of a general theory of the optimal design of transmission grids, so straightforward engineering calculations can rarely be decisive. In contrast, gas transmission may be overbuilt due to strong incentives derived from the market power of the pipelines, but the chances that insufficient capacity will be built to meet growing demand over the long term is remote.

Appendix E: Electricity Restructuring Facts & Figures

United States

California

History: The California region has been dominated by three large investor-owned utilities: Pacific Gas & Electric, San Diego Gas & Electric, and Southern California Edison, as well as a few sizeable municipal utilities such as the Los Angeles Department of Water and Power. The CA market has historically had electric rates some 30-50% higher than the national average. CA generation includes coal, oil, natural gas, nuclear, hydro, and geothermal resources, as well as imports from the Pacific Northwest, Arizona, and occasionally Mexico.

http://www.cpuc.ca.gov/electric_restructuring/restruc_plain.htm

System Conditions: The "Big Three" IOUs produce ~70% of California electricity. 1996 net summer capability was 43,342 MW. Total sales to ultimate consumers were 218 billion kwh with revenue of $20.67 billion. The CA ISO controls 75% of the power grid and can meet up to 45 GW of peak demand. Load forecast and transmission line information is available.

http://www.caiso.com/iso/news/FYIFAQs.html

Daily price information is available at the PX Website.

http://www.calpx.com/MarketPrices/DayAhead/dailyprice.html

Market Structure:

Oversight: The CA market is divided into two sections. The Independent System Operator (ISO) controls the high-voltage transmission grid and maintains reliability by exerting real-time control and contracting for ancillary services. The utilities retain ownership of the physical grid.

http://www.caiso.com/iso/news/FYIControlCenter.html

Market Separation: The ISO runs three separate markets: a real-time Imbalance Market, an Ancillary Services Market (day-ahead and hour-ahead; automatic regulation, spinning reserves, non-spinning reserves, and replacement reserves), and a Congestion Management Market ("adjustment bids" day- and hour-ahead). The PX runs a day-ahead and an hour-ahead market for energy.

http://www.caiso.com/iso/news/FYIMarkets.html
http://www.calpx.com/aboutpx.htm

Description: The ISO never takes a net position for itself, acting as a clearinghouse but never buying or selling power itself. The three big IOUs are required to sell their generation to and buy power from the PX through March 2002. Retail competition began March 31, 1998.

http://www.caiso.com/iso/news/FYIMarkets.html

Market Processes:

Bid Structure: Scheduling Coordinators submit at 10am, response from ISO at 11am, revised schedules at noon, markets close and congestion charges calculated at 1pm. The ISO takes increment/decrement bids for congestion relief in a completely separate market from energy bids.

http://www.caiso.com/iso/news/FYITimeTable.html

Tariffs: A mandatory 10% rate reduction is financed by a *bond charge* through 2002. A *competitive transition charge* finances the accelerated recovery of stranded costs by the utilities. A *public goods charge* funds research and development into new technologies and conservation programs.

http://www.cpuc.ca.gov/electric_restructuring/restruc_plain.htm

Settlement: All energy sold in each period has the same price. Payments to generators are based on the marginal producer for that period. Contracts made in one market are binding at the price made in that market.

Planned Developments: Consumer choice will be phased in gradually over the next few years.

New England (ISO-NE)

History: The New England states do not have an overall structured plan yet; however, most states seem to be moving forward with various proposals of their own. NEPOOL may or may not prove to be the overarching governing body.

System Conditions: As of 8:00am July 8th, installed generating capacity was 21,109 MW for NEPOOL. Peak load for July 7 was 16,649 MW at 2pm. ISO-NE has replaced the old NEPOOL.

http://www.iso-ne.com/power_system/morning_report_external.html

Northeastern Maine is more connected to Canada than the rest of the U.S.; there is a fear that deregulation may hurt rural customers more than it will help them, because of "cherry-picking" by large generators contracting only with large businesses.

http://www.emec.com/deregulation/

Market Structure:

Oversight: Information about the ISO--NEPOOL agreement can be found at

http://www.iso-ne.com/about_the_iso/organizational_structure.html.

Description of its Board of Governors is at

http://www.iso-ne.com/about_the_iso/iso_board.html.

Market Separation: The DPU Restructuring Proposal outline in Massachusetts envisions an ISO and a power exchange (PE) independent of the utilities. Electric utilities would separate into distinct corporations with rules governing inter-affiliate transactions. They aim for a competitive generation market by January 1, 1998.

http://www.energyonline.com/Restructuring/models/dpu.html

Vermont also advocates creation of an ISO and a PX, but doubts that NEPOOL could achieve sufficient independence. The Board seems to favor several independent non-profit PXs, functioning within and across regions. Utility divestiture is “appropriate but not required.”

http://www.energyonline.com/Restructuring/models/vt_draft.html

The Maine plan requires IOUs to divest by March 1, 2000, which would also be the start date for retail competition.

http://www.energyonline.com/Restructuring/models/mainelaw.html

ISO-NE is a "residual" wholesale electricity market--to the extent a participant generates electricity in excess of the demand of its customers, the remainder goes into the pool.

http://www.iso-ne.com/about_the_iso/

Description: The Mass. DPU proposes a price-cap system of performance-based regulation.

http://www.energyonline.com/Restructuring/models/dpu.html

Maine would require every electric provider to have at least 30% renewables with tradable credits.

http://www.energyonline.com/Restructuring/models/mainelaw.html

UI advocates method of transmission pricing based on generator impact on system flows. The spread of power flows is measured as the difference in MW flow observed on each line in the region when the results of two computer simulations are compared--one with the generator on at full output, and the other with the generator off. This recognizes that the transmission system has to be able to accommodate the full output of the generator. It is important to note that the spread pattern will be constant, regardless of what else is happening throughout the system, as long as the transmission system configuration is not changed.

The MW impact on each line is multiplied by that line's length to get a MW-mile value. These values for all lines in the region are summed to arrive at a total MW-mile value for the generator. Adjustments to the MW-mile value are made to recognize line cost differentials, transformers, and transmission interface restrictions or congestion. Each generator's payment responsibility is the total regional transmission costs multiplied by a fraction; the fraction equals that generator's MW-mile value divided by the sum of the MW-mile values for all the generators. The generators pay these amounts into a regional fund. For a generator, this payment becomes a cost component similar to a fixed charge cost component to be recovered in its sales of power, whether to native load customers or to wholesale buyers. The regional transmission fund receipts are distributed to all transmission owners on the basis of their actual costs. No other transmission arrangements are needed to engage in power transactions.

http://www.energyonline.com/Restructuring/models/uiprop2.html

Market Processes:
Mass. DPU proposes a regional, zoned network transmission tariff. It mandates a 10% rate cut beginning March 1, 1998 with an additional 5% on September 1, 1999.

http://www.energyonline.com/Restructuring/models/dpu.html

The Vermont Public Services Board believes its plan will "create the opportunity for full recovery of stranded costs provided they are legitimate, verifiable, otherwise recoverable, prudently incurred, and non-mitigable." Unmitigated stranded costs would be recovered through a utility-specific "Competition Transition Charge," or CTC. Each utility's CTC would be imposed on retail customers using the state's transmission and distribution system and be collected through electric bills. The CTC would be non-bypassable, competitively neutral and applied to both present and future customers.

http://www.energyonline.com/Restructuring/models/vt_draft.html

Connecticut has passed restructuring legislation in mid-April 1998 calling for consumer choice by the year 2000 and mandating a 10% rate cut until 2002.

http://www.deregulation.com/mapelectric.htm

ISO New England is a "day-ahead-hourly" market. Transmission pricing is a uniform flat rate, reviewed and approved by FERC.

http://www.iso-ne.com/about_the_iso/

New York

History: "In the 1994 New York State Energy Plan, the three state agencies (the NYS Public Service Commission, the NYS Energy Office and the NYS Department of Environmental Conservation) comprising the State Energy Planning Board identified several key factors contributing to New York's high electric rates. The leading factors include higher utility taxes, higher utility wages and benefits, higher utility operation and maintenance costs, and higher capital costs..." Also, the document below claims that IPPs have been the reason that deregulation is being considered—that they've proven that electricity can be produced more cheaply and efficiently.

http://www.ippny.org/factshee.htm

System Conditions: New York's 1996 net summer capability was 32,112 MW. In October 1994, the 1994 New York State Energy Plan states that on average New York's electric rates are nearly four cents higher than the national average rate of 7.06 cents per kilowatt-hour.

http://www.ippny.org/factshee.htm

Market Structure:

Market Separation: Under the IPPNY model, a fully competitive statewide market for wholesale power would be formed, overseen by an Independent System Operator (ISO) whose sole job would be to maintain the current high levels of system reliability while coordinating transfers of bulk power from buyers to sellers throughout the state. In the routine performance of this job, the ISO would lease control of all transmission plants in the state and contract for whatever load following, voltage control and other ancillary services were necessary from the generation (and possibly demand-side) markets on a least-cost basis.

http://www.ippny.org/cob_summ.htm

Description: Bilateral, official PX, and unofficial arising competitors are all allowed and encouraged. IPPNY is strongly opposed to any sort of government-mandated centralized pool into which generators must sell, or from which customers must purchase. The matching of output and consumption, and the establishment of a market clearing price are not monopoly functions. Monopolies and government regulators therefore have no business either mandating or regulating entities established to serve those purposes. IPPNY and many other parties believe that "…a mandatory PoolCo would be the antithesis of competition and represent a giant step backward toward the central regulation of markets that has failed everywhere it has been attempted."

http://www.ippny.org/cob_summ.htm

Many New York State consumers are now starting to choose who will supply their electricity, natural gas, and other services instead of the local utility company. They may select to make arrangements through either an energy services company (ESCO) or marketer. Or, they may choose to have an agent to serve as their intermediary between the marketer and the local utility company. Various utilities open up their areas on various time schedules.

http://www.dps.state.ny.us/cobelec.htm
http://www.dps.state.ny.us/escoalert.htm

Market Processes:

Tariffs: All associated stranded costs should be recovered, for both utilities & IPP's...to prevent utilities from recovering stranded costs would be "neither ethical nor prudent."

http://www.ippny.org/cob_summ.htm

The IPP New York plan expects a 3-4 c/kWh price and recommends imposing a non-bypassable access fee to all customers to recover utility & IPP stranded costs. A Systems Benefits Charge (SBC) should be levied to fund investments in renewables, low-income customer assistance, etc.

http://www.energyonline.com/Restructuring/models/ippny3.html

Pennsylvania—New Jersey—Maryland (PJM)

History: Established in 1927, The PJM service area includes all or part of Pennsylvania, New Jersey, Maryland, Delaware, Virginia, and the District of Columbia.
http://www.pjm.com/about/general.html

System Conditions: PJM today handles almost 8% of the country's electric power, with a pool generating capacity of over 56,000 MW and a 1997 peak load of 49,409 MW.
http://www.pjm.com/about/general.html
Daily market and system data is available at the following Website.
http://www.pjm.com/account/control_page.html

Market Structure:

Oversight: PJM is governed by an independent Board of Managers. The Board appoints the PJM President and CEO to direct and manage PJM's day-to-day operations. Contractual agreements define the specific relationships and services that PJM provides to both its Energy Market and Transmission Service Participants. The PJM Members Committee, comprised of representatives of PJM members, advises the Board of Managers.
http://www.pjm.com/about/general.html
PJM Interconnection became the first operational Independent System Operator in the U.S. on January 1, 1998, managing the PJM Open Access Transmission Tariff and facilitating the Mid-Atlantic Spot Market There is no separation of the PX and ISO; it is a tight centrally-dispatched pooling system.
http://www.caiso.com/iso/news/FYIControlCenter.html

Market Separation: The PJM staff centrally forecasts, schedules, and coordinates the operation of generating units, bilateral transactions, and the spot energy market to meet load requirements. To maintain a reliable and secure electric system, PJM monitors, evaluates and coordinates the operation of over 8,000 miles of high-voltage transmission lines. The PJM OASIS is used to reserve transmission service. Operations are closely coordinated with neighboring control areas, and information is exchanged to enable real-time security assessments of the transmission grid. PJM provides accounting services for energy, ancillary services, transmission services, and capacity reserve obligations.
http://www.pjm.com/about/general.html

Description: On April 1, 1998, PJM started the first energy market based on Locational Marginal Pricing which takes into account both generation marginal cost and the generator's location on the physical system. Prices are based on system conditions at the time. LMP's include transmission congestion cost, and the ISO calculates prices at 5-minute intervals. Includes 1750 PJM busses plus some outside; prices calculated individually for all of them as well as for aggregated groups. They claim that "Transmission limitations in neighboring control areas will NOT affect LMP's in the PJM area."
http://www.pjm.com/lmp/info/control_page.html
http://www.pjm.com/ferc/filings/19971231/lmpfaq.pdf

Firm Transmission Rights (FTRs) have had an initial allocation, plus there is a secondary market for trades. All trades must be simultaneously feasible under normal operating conditions. Assignment is point-to-point or generation capacity area-to-aggregate load area. "Network Service FTRs" are recalculated on an annual basis, while "Point-to-Point FTRs" can be changed on annual, monthly, weekly, or daily bases depending on the particular reservation period. FTRs are a financial obligation and may become a liability when the path is opposite actual congested flow, so charges are possible as well as credits. All FTRs are posted on PJM's OASIS system. An FTR is credited as the LMP differential between the designated source and sink of the FTR, multiplied by the MW value.

http://www.pjm.com/lmp/info/control_page.html

Market Processes:

Bid Structure: Currently, all generators are centrally dispatched at cost and are not using self-determined prices (FERC will be ruling on this shortly). Thus all resource bids are capped at cost. All interchange energy market transactions must be submitted day-ahead.

http://www.pjm.com/about/general.html

*Tariffs: F*or the open access transmission tariff see

http://www.pjm.com/ferc/filings/19980130/transtar.pdf

Settlement: FTRs completely hedge against congestion charges under normal conditions; if not at some point, they will be trued up at end of month from excess congestion charges (over-collection).

http://www.pjm.com/lmp/info/control_page.html

Other States and some General Resources

Alabama: 5/6/96---State Legislature & Governor passed bill to protect utilities from loss of the opportunity to recover the costs of investments made to serve customers when one of those customers left a utility's system to purchase electricity on a private contract. ~45 day waiting period, written notices required, judicial review.

http://www.energyonline.com/Restructuring/models/alabama.html

Arizona: 11/15/97 -- ISO Working Group releases conclusions on Desert STAR (Southwest Transmission and Reliability Operator). An Arizona-only ISO is infeasible; STAR covers Arizona, New Mexico, southern Nevada, and west Texas. Multi-state policies would be specifically designed to eliminate anti-competitive effects of pancaked rates. No formal PX is needed for the SW region. May implement an Independent Transmission Operator in AZ to facilitate changes of open retail access and non-discriminatory transmission pricing. Details responsibilities and powers of a hypothetical ISO.

http://www.cc.state.az.us/working/iso.htm

Federal level: The Clinton Administration released a plan for electric utility restructuring on March 28, 1998. Five main objectives: (1) encouraging States to implement retail competition; (2) protecting consumers by facilitating competitive markets; (3) assuring access to and reliability of the transmission system; (4) promoting and preserving public benefits; and (5) amending existing federal statutes to clarify federal and state authority. Mandates retail-level choice by Jan 1, 2003.

http://www.energyonline.com/Restructuring/models/white01.html

Generating capability (refs):

http://www.eia.doe.gov/cneaf/electricity/epa/epav1t4.dat
http://www.eia.doe.gov/cneaf/electricity/ipp/t17p01.txt

Sales to consumers, consumption and revenue data:

http://www.eia.doe.gov/cneaf/electricity/esr/t06.txt

Restructuring links, state-by-state:

http://www.ai.org/iurc/electric/table97c.html
http://www.nrri.ohio-state.edu/restruct/states.html

An argument against stranded cost recovery on the part of utilities:

http://www.heritage.org/library/categories/regulation/tp20.html

Other Countries

Argentina

History: The Argentine electricity privatization effort began in 1992. By this time, Argentina's electricity industry had deteriorated badly and was characterized by severe operational and financial difficulties. The industry was constantly threatened with the possibility of blackouts, a threat which worsened during periods of relatively little rainfall (such as the summer) because of Argentina's reliance on hydroelectric power generation. Electricity was also expensive and often stolen by consumers either through illegal hook-ups or by failure to pay bills. Post-privatization, the rate of growth of the industry leaped from 3%/yr to 7%/yr. Argentina modeled after Chile's deregulation efforts, but tried to improve in some parts (i.e., they required complete separation of transmission from distribution, and ruled that no single generating company can provide more than 10% of the national capacity). About 80% of country's generating capacity has been privatized.

http://www.eia.doe.gov/emeu/pgem/electric/ch4l1.html

System Conditions: The post-privatization Argentine power generation industry (including both conventional and non-conventional power facilities) is composed of independent, largely unregulated power generation companies. The companies are essentially unregulated because electric power generation is considered a competitive market. The nearly 40 generating companies operating in Argentina are assured by the national electricity regulatory body (Enre) of having open and equal access to the national grid and receive unregulated prices. In order to avoid market concentration difficulties, generation companies are legally restricted to a market share of 10 percent or less of the national electricity sales volume. As of May 1996, the base price paid for electricity dispatched by Cammesa was $5 per megawatt hour and the fee paid for reserve capacity also was $5 per megawatt hour, summing to $10 per megawatt hour.

http://www.eia.doe.gov/emeu/pgem/electric/ch4l4.html

Market Structure:

Market Separation: The wholesale electricity market (also known as a power pool) has both a supply side and a demand side. The interaction of the supply and demand sides of the wholesale market largely determines wholesale prices for electricity. Three kinds of wholesale electricity prices exist in the Argentine electricity industry: contractual prices, seasonal prices, and spot prices. Of these, seasonal and spot prices are determined directly in the wholesale market, while contractual prices are affected indirectly by the wholesale market.

http://www.eia.doe.gov/emeu/pgem/electric/ch4l4.html

Description: As in the United Kingdom and Australia, electricity transmission has been defined by Enre (the national electricity regulator) as a natural monopoly and is closely regulated. Firms may enter the industry only after successfully bidding for a fixed-duration concession for a particular area and may charge no more than regulated prices for their services. Concessionaires are required to allow open access to their transmission network to third parties. Transmission companies are not allowed to buy or sell electricity. Instead, their revenues come exclusively from the regulated prices they receive. The price is based on the availability (providing a fixed source of income) and the use (providing a variable source of income) of their network assets. The rate at which they are paid is capped by the federal electricity regulatory body, providing an incentive for Argentine transmission companies to reduce their costs. More than half of the transmission companies have been at least partially privatized. Distribution rates are also capped by RPI-X.. Price-cap regulation is applied to transmission and distribution but not to generation. Large users (more than 4,380 MW annual consumption) can choose to be directly supplied by a generation company, in which case they enter into bilateral negotiation. They may also be supplied by a distribution company at the normal rate. Finally, they may buy directly from the market at the spot price. Contract length is typically one year and is largely unregulated. The government sets some seasonal prices based on weather conditions and water level.

http://www.eia.doe.gov/emeu/pgem/electric/ch4l4.html

Market Processes: A fixed charge is added to all of the market-determined prices to cover payments made by Cammesa to power generators providing reserve capacity to the electricity grid. Transmission loss charges are based on the physical distance of the electricity seller from Buenos Aires. The greater the distance, the more the price received is discounted to cover transmission losses. Finally, the Argentine restructuring has been a marked success, leading to prices about 40% lower and markedly increased reliability. Productivity and foreign investment have increased dramatically, with inflation dropping.

http://www.eia.doe.gov/emeu/pgem/electric/ch4l4.html

Australia

History: Prior to recent reforms, electricity supply in Australia was provided by vertically-integrated, publicly-owned state utilities with little interstate grid connections or trade. The Australian national electricity market (NEM) is to develop in stages until a fully competitive market for electrical generation and retail supply is achieved by 2001. Although it is referred to as a national market, NEM will initially include the states of Victoria, New South Wales (NSW), South Australia, Queensland, and the Australian Capital Territory, with the possibility of an expansion into Tasmania following its grid interconnection. Western Australia and the Northern Territory will not participate in the market due to geographical and cost factors—basically, they're too far away with too little population.

http://www.eia.doe.gov/emeu/pgem/electric/ch3l1.html

System Conditions: Australia runs about 80% coal-fired generation. Most consumption is in the eastern states of NSW (35%), Victoria (23%), and Queensland. Tasmania is the exception, the island mostly using hydropower.

http://www.eia.doe.gov/emeu/pgem/electric/ch3intro.html

Electricity transfers from Victoria to NSW during off-peak periods have ranged between 500 and 1,100 MW. NSW generators have taken on an expanded role in the intermediate/peak periods as a result of their greater flexibility, commonly exporting 100 to 500 MW to Victoria during these peak periods.

http://www.doe.nsw.gov.au/Electricity/EnergyReform/infosheet3.html

Daily reports of the Queensland Electricity Market Operations can be found at http://www.nemmco.com.au/NEM_Web/NEM_Resources/Market_Data/Qld_Interim/Qldinterim.htm

Market Structure:

Oversight: Most Australian states are in various stages of deregulation, with Victoria being the most advanced, having privatized most of its generation and distribution facilities. Other states have unbundled utilities into separate entities but have not sold them to private investors yet.

http://www.eia.doe.gov/emeu/pgem/electric/ch3l4.html

New South Wales began restructuring efforts beginning in 1995. First, it aggregated 25 distribution businesses into six, and separated within those businesses the monopoly network functions from contestable retail services. Then, it transferred the transmission functions of Pacific Power to a separate organization, the Electricity Transmission Authority, trading as TransGrid. Finally, it transferred generation assets to create three competing businesses: Pacific Power, Macquarie Generation and Delta Electricity. In March 1996, all newly created distribution and generation businesses (except Pacific Power) were established as State Owned Corporations. Competition begin in May 1996.

TransGrid is responsible for development, operation, and administration of the market, and for power system security. Retail competition began in 10/96.

http://www.doe.nsw.gov.au/Electricity/EnergyReform/infosheet3.html

Market Separation: The National Electricity Code ("the Code") establishes the regulatory and operational framework of the new Australian national electricity market and binds all participants in the wholesale power generation market to the specified rules. The Code addresses the following: market rules; grid connection and access; metering; network pricing (transmission and distribution); system security, and procedures for Code administration.

http://www.eia.doe.gov/emeu/pgem/electric/ch3l2.html

Contestable customers have two options: (1) participate in the wholesale market, or (2) participate in the retail market. If the contestable customer decides to participate in the retail market, all of their electricity will be supplied through a marketer and they cannot participate in the wholesale market. On the other hand, if the contestable customer chooses to trade in the wholesale market they must register as a participant with the NEMMCO. In the wholesale trading market there will be three levels of trading: via a long-term bilateral contract; via a short-term forward market; and via a spot trading market. Participants in the wholesale market can operate in any combination of these markets. In the wholesale environment, electricity buyers could be both end-use customers as well as marketers. Sellers of electricity could be generators as well as marketers. Short- and long-term contracts are both used. Contestable customers may negotiate retail contracts, thereby opting out of the pool (perhaps due to risk aversion).

http://www.eia.doe.gov/emeu/pgem/electric/ch3l3.html

Description: Since NEM1 commenced pool prices in NSW have fallen from an average of around $23 per MWh to around $10-13 per MWh. However, transitional contracts put in place between the generators and retailers by the Government (vesting contracts) have protected generators from the full impact of price falls.

http://www.doe.nsw.gov.au/Electricity/EnergyReform/infosheet3.html

Demand is forecast and interstate trading is determined by the "interconnection scheduling module." A price cap of $5/kwh may be set, along with use of a CPI-X formula very similar to the U.K.'s RPI-X.

http://www.eia.doe.gov/emeu/pgem/electric/ch3l5.html.

The role of NEMMCO is to manage the wholesale electricity market on behalf of the participants. It will receive bids from the participants and settle the short-term forward market and spot market. It will calculate the dispatch order for the physical operation of the system and pass this to the System Operator. Operates on a self-funded basis. Manages and settles the Pool.

http://www.nemmco.com.au/nemmcorol.htm

The Australian electricity market is currently in a "Transitional Phase" where states operate under a separate wholesale power pool but generators compete directly with each

other. The national government is currently engaged in building three large-capacity links to connect the intrastate grids; in 1995, interstate flows of electricity represented on average less than 2% of total electricity in SE Australia.
http://www.eia.doe.gov/emeu/pgem/electric/ch3l1.html

Market Processes: The dispatch is based on an LP optimization code. The following URL contains discussions of how the spot price will be set in critical situations.
http://www.nemmco.com.au/markdesign.htm
Interregional transfer capabilities in Australia are based on transient stability, with analysis of multiple situations; regression on variables. Marginal loss factor calculations, ancillary services charges, etc.
http://www.nemmco.com.au/mip_operate1.htm

Planned Developments: Details of the planned Australian Power Pool can be found at http://www.eia.doe.gov/emeu/pgem/electric/ch3l5.html.

New South Wales issued a revised timetable for electricity reform. 1) Contestability for customers in the range of 160 Megawatt hours per annum--750 MWh pa will proceed as previously announced from 1 July 1998. 2) Aggregation for customers below 160 MWh will be permitted from 1 July 1999 subject to certain criteria including a minimum site threshold of 100 MWh pa. 3) Retail contestability for all remaining customers will commence 1 January 2001 with detailed transitional arrangements to be developed and announced at a later date. They say their plan "...stands in stark contrast to Victoria, which gives small business no options for aggregation."
http://www.doe.nsw.gov.au/Electricity/Timetable/RETMTBLE.html

Brazil

Again, little useful information has been found about electricity reform efforts in this country. Some system statistics may be useful: 94% of Brazilian electricity comes from hydroelectric sources. One large utility (Electrobras) controls 60% of the market; the remainder comes from small power producers. The document states that privatization of federal generation companies will take place in mid-1997. Brazil had a total generation of 256 GWh in 1994 and 55 GW of installed capacity as of June 1995. (Tables at the below URL categorize generation information by utility, consumer type, etc.)
http://www.energyonline.com/Restructuring/models/brazil1.html

Canada--Alberta

System Conditions: Pool price, demand and supply figures, unit generation status, etc. are available at the URL below. It seems that Alberta only has three thermal generators and one hydro generator, which may make it good for modeling purposes.
http://www.powerpool.ab.ca/

The Alberta consumptive market is 50,000,000 MWh of generation per year, varying from 5000 to 7500 MW per hour. Total value is $2 billion CA.

http://www.electricity-markets.com/whitepaper/SheldonFulton/newforwards_frame.htm

At present, 95% of Alberta electricity is produced by three companies. Market power studies are underway.

http://www.energy.gov.ab.ca/electric/restruct/overview.htm

Current pool price information, system conditions, etc. can be found at

http://www.powerpool.ab.ca/.

Market Structure:

Oversight: Oversight and governance details of the Alberta power pool can be found at

http://www.powerpool.ab.ca/the_pool/the_pool.html.

Market Separation: A proposed futures market separates physical volume from "business" volume with balancing calculated ex post.
http://www.electricity-markets.com/whitepaper/SheldonFulton/newforwards_frame.htm

Description: 85% of the Alberta market is price protected by legislative hedges with the remaining 15% available for trade in the forward market. They envision several types of contracts:

For the spot market,

Base-load 24 hour continuous flow
Mid-peak w/fixed hourly deliverability
On-peak 4-hour blocks

Near-term forward contracts (month-long)
Longer term forward/futures/derivatives

Eventually maybe futures, options, or multi-year bilateral contracts will be offered.
http://www.electricity-markets.com/whitepaper/SheldonFulton/newforwards_frame.htm

Legislated price hedges have been successful in restraining market power; however, they might also be distorting the pool price. Designers hope that long-term CFDs might mitigate market power by the three dominant incumbents; threat of entry in the generation market is low.

http://www.energy.gov.ab.ca/electric/restruct/mrkpwr3.htm

Market Processes:

Tariffs: Residual benefits & stranded costs are currently measured and shared on an ongoing basis by means of legislated financial arrangements (hedges) set up January 1, 1996 under the Electric Utilities Act.

http://www.energy.gov.ab.ca/electric/restruct/overview.htm

The description of the proposed Alberta Open Access Power Pool includes provisions for bilateral CFDs and a “postage stamp” transmission charge.

http://www.energyonline.com/Restructuring/models/alberta.html

Planned Developments: The regulated arrangements (hedges) will be eliminated no later than January 1, 2001 and replaced with a tradable instrument, such as long-term tradable contracts, to ensure the fair sharing of costs and benefits. Customer choice will start in July 1999 at the retail level; power services will be unbundled then as well.

http://www.energy.gov.ab.ca/electric/restruct/overview.htm

Canada--Ontario

Market Structure:

Oversight: First reading of Ontario's Energy Competition Act on 6/9/98. Separates Ontario Hydro G off from T&D, establishes non-profit crown corporation called Independent Electricity Market Operator to ensure reliability and fair access.

http://www.omdc.org/NewHomePage.html

Market Separation: The Independent Market Operator (IMO) will perform least-cost dispatch; operating both the PX and the grid. Will be both day-ahead, hour-ahead, real-time imbalance bidding, as well as bilateral contracts for consumers.

http://www.omdc.org/ArchivedDocuments.html

Description: Suppliers and consumers can choose either the spot market or bilateral contracting. The IMO will administer tradable transmission rights, allocated initially to load-serving entities and periodically auctioned off. Ancillary services secured through day-ahead markets. Charges for real-time balancing (inc/dec) costs are assumed to be the same as participants' submitted bids and are not separately submitted. Both bilateral and PoolCo will be accommodated. Both physical contracts and CFDs will be allowed, as long as appropriate rules to prevent cost shifting are employed.

http://www.omdc.org/ArchivedDocuments.html

Market Processes:

Tariffs: A uniform region-wide access fee will be charged to all loads to recover fixed costs of transmission.

http://www.omdc.org/ArchivedDocuments.html

Planned Developments: Many archived documents about market design issues being examined at the site below; many details about what they've investigated, but not so many on what will be adopted. Goal is to have competition by 2000.

http://www.omdc.org/NewHomePage.html

Demand-side bidding may or may not be used. Debating whether to employ just one market or multiple time-ahead markets with different settlement prices.

http://www.omdc.org/ArchivedDocuments.html

New Zealand

Little information about electricity restructuring in New Zealand has been found so far. Some price figures and data can be found at

http://www.transpower.co.nz/grid/February_1998/Wholesale_electricit.shtml.

For example, in January 1998 (the latest month for which figures are available) the average wholesale electricity price was 4.10 c/kwhr on the North Island and 4.36 c/kwhr on the South Island.

Scandinavia (NordPool)

History: Norway deregulated in 1991, although it retained public ownership of the utility companies. It combined with Sweden in 1996 to form NordPool.

http://ksgwww.harvard.edu/hepg/FPovertext.html

System Conditions: Production, consumption, reservoir content, day-ahead and spot price information can be found at

http://193.69.80.130/marketinfo/index.html.

Market Structure:

Market Separation: The price mechanism in the spot market is used for regulating the power flow in situations of limited capacity in the central grid. The market can be considered a combined energy and capacity market. The price calculation is based on the balance price between the supply and demand for all the participants.

http://193.69.80.130/menu1/m1_3.htm

Description: The generating companies are not required to sell through the PX. The PX merely handles the spot market.

http://ksgwww.harvard.edu/hepg/FPovertext.html

Since the establishment of the futures market there has been a change in the use of the market from physical trading to risk management in power portfolios. During this period NordPool has further developed the market to meet increasing demand. Among other things, new trading mechanisms have been developed with the introduction of telephone-operated trade in the Market Equilibrium. This development increases the need for information, availability and trading and settlement mechanisms as well as product capacity. As a consequence the exchange has developed the futures market into a financial market.

http://193.69.80.130/menu1/m1_4.htm

Most generation in NordPool is covered by futures contracts.

http://ksgwww.harvard.edu/hepg/FPovertext.html

Market Processes:

Tariffs: The choice of method for treating bottleneck situations is different in Norway and Sweden. In Sweden, Svenska Krafnat makes use of the counter purchase principle.

This means that Svenska Kraftnat pays for the downwards regulation of the surplus area and for the upwards regulation in the deficit area. The costs connected with the counter purchase is regained through tariffs for power transmission. In the Norwegian system the price mechanism is used. The price is reduced in the surplus area and increased in the deficit area until the transmission need is reduced to the capacity limit. This means that the market participants are charged the costs through the capacity fee in the market settlement. In the spot market the price mechanism is used to handle bottlenecks between reporting areas with Sweden always as one and Norway as one or several reporting areas.

http://193.69.80.130/menu1/m1_3.htm

Settlement: Futures contracts can be divided into two main categories, futures and forwards. The products traded on the futures market are futures contracts. The difference between the two types of contracts lies in the settlement of contracts trading period, i.e. the due day (delivery week). For futures contracts the value of each participant's contract portfolio is calculated daily derived from a change in the market of the contract. The daily evaluation will be settled economically between the buyer and the seller. In this way contract losses are quickly exposed and realized at the same time as profits are realized and paid out.

http://193.69.80.130/menu1/m1_4.htm

Spain

History: Spain restructured its electricity industry with the Electricity Acts of 1994 and 1997. Open transmission access was granted, and complete unbundling is planned by 31/12/2000.

http://www.csen.es/pdf/FAQ.pdf

System Conditions: Spain's 1997 net production was 166,908 GWh, divided into approximately 33,138 GWh of hydro, 55,305 of nuclear, 62,098 of coal, 6,634 of natural gas, and 209 of oil. Total net export capacity was 7,710 MW (to France, Portugal, and Morocco), limited to 2,000 MW for network reliability reasons. Total demand was 162,017 GWh with a maximum load of 27.4 GW (load factor 0.7). In total there is 44,000 MW of installed capacity with a high degree of horizontal integration. The division of operating revenue among utilities is given on page 23, and the division of generation capability is on p. 24 of the below URL.

http://www.csen.es/pdf/FAQ.pdf

A total of 51,000 GWh were traded over the first four months of 1998 on the day-ahead market, average prices of 4.3 pesetas/kWh with little price fluctuation. Prices paid for different generation technologies ranged from 4.201 Ptas/kWh for nuclear plant to 5.503 Ptas/kWh for pumping unit turbine power. Coal power plant obtained much higher prices than hydro power generation due to the peak modulation role played by conventional thermal power plant during the market's first 1.5 months of operations. 189 GWh were traded on the intra-day spot market in April, 1998. Weighted price average was 3.092 Ptas/kwh, with a high of 6.6 and a low of 0. The impact of restrictions and second-tier regulation on electricity purchasing price was 0.139 Ptas/kWh; with an unknown effect of third tier regulation and deviation management.

http://www.csen.es/hojas/82e.htm

Market Structure:

Oversight: All electrical utilities will privatized by June, 1998. The state owns 25% of Red Electrica de Espana (REE), the grid operator.
http://www.csen.es/pdf/FAQ.pdf

Market Separation: Daily markets and markets for secondary and tertiary operations (ancillary services) commenced January 1, 1998. An intra-daily market began on April 1, 1998. There is a mandatory bid regime for units whose installed capacity is greater than 50 MW. A market for physical bilateral contracts is also in place, with a retail market developing.
http://www.csen.es/hojas/82e.htm
The wholesale market is governed by a System Operator (in charge of security, reliability, and technical management) and a Market Operator (who settles transactions, matches operations, and settles payments between agents). Bilateral physical contracts are allowed and coexist with the wholesale market. A futures market may be developed in the future, if warranted.
http://www.csen.es/pdf/FAQ.pdf

Description: The sequence of operations for a particular day begins with the day-ahead market where the basic schedule and prices per time interval are obtained for the following day. Once possible technical restrictions have been analyzed, it may be necessary to modify the basic schedule in line with the order of priority stipulated by the Market Operator and with the bids that have been submitted to deal with the restrictions. The next step is the convening of the ancillary services markets by the System Operator who notifies the agents about the second and third tier regulation requirements at this stage. This sequence of events results in the definitive feasible daily schedule. Once the definitive feasible daily schedule is known, the first intra-day spot market session can be opened. The outcome of this session is the basic intra-day schedule which subsequently becomes the final scheduling when any possible modifications prompted by technical restrictions have been added into it. Once the intra-day market has been fully implemented, there will be twenty-four sessions a day, one for each hour. During the initial, transitional period, however, starting April 1st and ending on June 30th 1998, the market is being phased in with only two sessions a day. The actual system operation is carried out by REE through the implementation of the final scheduling and the use of the ancillary services that are contracted for. Should deviations exceed a certain pre-set level, REE convenes a deviation management session whose purpose is to cut the additional cost that might be entailed by operating with only the statutory ancillary services. REE can also use exceptional albeit non-emergency procedures in situations where the ancillary services are not sufficient to guarantee the required security and reliability of operations.
http://www.csen.es/hojas/82e.htm

Market Processes:

Bid Structure: All markets work in a similar way. Bids are sent electronically to the relevant market operator's computer by the agents. The operator matches the bids, *i.e.* it decides on the accepted price and power level for each bid. The operator then notifies all the agents of the outcome of the matching process.

http://www.csen.es/hojas/82e.htm

Tariffs: The tariff on production is based on the average wholesale market price of electricity corresponding to the customer's profile. An extra 5% tax funds research into the diversification of Spain's energy resources. Consumers with more than 15 GWh annual consumption are qualified customers and may purchase on a regulated tariff.

http://www.csen.es/pdf/FAQ.pdf

United Kingdom

History: The United Kingdom implemented the first, largest, and most heavily criticized electricity restructuring. Electricity privatization occurred in the larger context of the privatization of many formerly state-owned U.K. industries during the 1980's under Margaret Thatcher. The old British electricity industry was primarily dependent on British coal and nuclear power, both of which were more costly to run than other forms of generation. A detailed history of old British electricity regulation and initial deregulation steps can be found at

http://www.eia.doe.gov/emeu/pgem/electric/ch2l1.html

On Vesting Day, April 1, 1990, private companies took charge of the industry. The Office of Electricity Regulation (OFFER) now regulates transmission and distribution under a price cap. Market-based pricing was intended; although OFFER does not set prices, it has considerable influence due to its ability to refer cases to the Monopolies and Mergers Commission (MMC). Concerns over whether the generation business was sufficiently competitive and was behaving as a duopoly have caused OFFER to intervene several times after privatization. A temporary price cap was in effect from mid-1994 through 1996. Some mergers and acquisitions have been blocked by regulators while others have been allowed. Reforms have been entirely top-down, by federal government mandate, rather than state-by-state experiments as in the U.S. or Australia.

Market Structure:

Oversight: Details of how shares in the various U.K. privatized companies were distributed are in

http://www.eia.doe.gov/emeu/pgem/electric/ch2l4.html.

Market Separation: Power is required to be sold to one national wholesale pool with one national price. Distribution is managed by the twelve RECs.

http://www.eia.doe.gov/emeu/pgem/electric/ch2l1.html

Description: Generators whose capacity exceeds 100 MW are required to submit their generation units to dispatch by the National Grid Company (NGC). The price cap for transmission is RPI-X; that is, the retail price index less expected future productivity.
http://www.eia.doe.gov/emeu/pgem/electric/ch2l1.html
"The UK experience clearly demonstrates that when incentives exist to exploit market power, elimination of vertical monopoly power is not enough to ensure competitive pricing, at least in the short run. Horizontal market power is enough to allow the generators to implement strategic decisions to control output and prices."
http://ksgwww.harvard.edu/hepg/FPovertext.html

Market Processes: Complete operating details of the England/Wales power pool are at http://www.eia.doe.gov/emeu/pgem/electric/ch2l2.html.

Appendix F: Bibliography*

[[Insert here, and then edit, bibliographies from Chao and from Wilson]]
[[Can Joel Singer be assigned to compile references from JTOR, ABI, Lexus-Nexus, and other sources? – if not then could the EPRI librarian?]]
[[Should this Bibliography be annotated?]]

DTI report of July 1998,
Material on NordPool:
Material on Victoria, Australia:
Material on Argentina, Chile, Spain, Canada-Alberta, Canada-Ontario, New Zealand
Key authors: Hung-po Chao, Stephen Peck, Paul Joskow, William Hogan, Richard Green, David Newbery, Frank Wolak, Derek Bunn, David Harbord, Lawrence Ruff, Stephen Rassenti, Vernon Smith, Variya and Wu, Shmuel Oren, James Bushnell, Severin Borenstein, Kenneth Binmore, Steven Stoft, Edward Kahn,
Key journals: Electricity Journal, Energy Journal, Restructuring Today, Journal of Regulatory Economics,
Key institutes: EPRI, EEI, NRRI, UC-Berkeley's Energy Institute, Harvard Kennedy School of Government's institute run by Hogan,
Key consultancies: Putnam Hayes and Bartlett (PHB – now Hagler Bailly), National Economic Research Associates (NERA), Charles River Associates (CRA), Tabors Caramanis Associates (TCA), Brattle Group, London Economics,
Key Organizations: NERC, WSCC, IEEE,
Key agencies: FERC, Department of Energy (DOE) and its subgroups (e.g., Energy Information Service?), California Energy Commission (CEC), state regulators [see NRRI website] in MA-MDPU, CA-CPUC, NY-NYPSC, PJM, Texas (Railroad Commission?), New ones beginning restructuring: Arizona, Virginia, Alabama, Oregon, Montana, midwest, Georgia?
Foreign Agencies: Alberta Department of Energy, Ontario Market Design Committee; VicPool, New SouthWales (NSW) Department of Energy, NEMMCO in Australia; New Zealand; Staftnet and Statkraft in Norway, Svenska … in Sweden; Offer and Department of Trade and Industry (DTI) in UK; the Spanish agency,

---

Mark Armstrong, Simon Cowan, and John Vickers, Regulatory Reform: Economic Analysis and the British Experience, MIT Press, 1994.

James Bushnell and Steven Stoft, "Electric Grid Investment under a Contract Network Regime," Journal of Regulatory Economics, 10:61-79, 1996.

Hung-po Chao and Hillard G. Huntington, Designing Competitive Electricity Markets, Kluwer Academic Press, 1998.

Hung-po Chao and Stephen Peck, “A Market Mechanism for Electric Power Transmission,” Journal of Regulatory Economics, 10(1), 1996.

Hung-po Chao and Stephen Peck, “An Institutional Design for an Electricity Contract Market with Central Dispatch,” Energy Journal, 18(1):85-110, 1997.

Michael Crew (ed.), Pricing and Regulatory Innovations under Increasing Returns, Kluwer Academic Press, 1996. [see article by Patrick and Wolak, "Industry Structure and Regulation in the England and Wales Electricity Market"]

Michael A. Einhorn (ed.), From Regulation to Competition: New Frontiers in Electricity Markets, Kluwer Academic Publishers, 1994. [see especially Chapters 3 by Vickers and Yarrow and 4 by Green]

Nils-Henrik von der Fehr and David Harbord, "Spot Market Competition in the U.K. Electricity Industry," Economic Journal, 103: 531-546, May 1993.

Richard Gilbert and Edward Kahn (eds.), International Comparisons of Electricity Regulation, Cambridge University Press, 1996.

Richard Green, "The Electricity Contract Market," Cambridge University, mimeo, May 1996.

Richard Green, "Increasing Competition in the British Electricity Spot Market,' Journal of Industrial Economics, 1997.

Richard J. Green and David M. Newbery, "Competition in the British Electricity Spot Market," Journal of Political Economy, 100(5): 929-953, October 1992.

Alex Henney and Simon Crisp, "Lessons for the U.S.? Transmission Pricing, Constraints, and Gaming in England and Wales," Electricity Journal, January-February 1997, 17-23.

William Hogan, “Contract Networks for Electric Power Transmission,” Journal of Regulatory Economics, 4:221-242, 1992.

S. Hunt and G. Shuttleworth, Competition and Choice in Electricity, John Wiley and Sons, 1996.

Paul Joskow, “Restructuring, Competition and Regulatory Reform in the U.S. Electricity Sector,” Journal of Economic Perspectives, 11(3):119-138, 1997.

Paul Joskow and Richard Schmalensee, Markets for Power, MIT Press, 1983.

G. McKerron and P. Pearson (eds.), The British Energy Experience: A Lesson or a Warning, Imperial College Press, London, 1996.

David Newbery, "Power Markets and Market Power," Energy Journal, 16(3): 41-66, 1995.

Shmuel Oren, “Economic Inefficiency of Passive Transmission Rights in Congested Electricity Systems with Competitive Generation,” Energy Journal, 18(1):63-84, 1997.

A. Powell, "Trading Forward in an Imperfect Market: The Case of Electricity in Britain," Economic Journal, 103: 444-453, March 1993.

Lawrence Ruff, “An Efficient Competitive Electricity Industry: Can the Vision Become Reality,” Electricity Journal, 10(1):8-16, 1997.

John Vickers and G.K. Yarrow, "The British Electricity Experiment," Economic Policy, 12:188-232, 1991.

Frank Wolak, “Market Design and Price Behavior in Restructured Electricity Markets: An International Comparison,” Energy Institute, University of California, Berkeley CA, 1997.

Frank Wolak and Robert Patrick, "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market," Stanford University, mimeo, 1996

Catherine Wolfram, "Strategic Bidding in a Multi-Unit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales," Harvard University, mimeo, January 1997.