ELECTRICITY MARKET RESTRUCTURING: REFORMS OF REFORMS

William W. Hogan

May 25, 2001

20th Annual Conference
Center for Research in Regulated Industries
Rutgers University
May 23-25, 2001

Center for Business and Government
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts 02138

CONTENTS

ELECTRICITY MARKET RESTRUCTURING: REFORMS OF REFORMS

William W. Hogan[1]
May 25, 2001

Electricity systems present complicated challenges for public policy. In many respects these challenges are similar to those in other network industries in providing a balance between regulation and markets, public investment and private risk taking, coordination and competition. As with other such industries, natural monopoly elements interact with potentially competitive services, but electricity has some unusual features that defy simple analogy to other network industries. Following a reversal of a long-term decline in real electricity prices, the last two decades of the twentieth century were for the United States a time of reform, reaction, and reforms of reforms in electricity systems, moving slowly towards greater reliance on competition and markets. Changing technology, new entrants in the generation market, and a legislative mandate to provide access to the essential transmission facility accelerated a process that required major innovations in institutions and operations. Complete laissez-faire competition is not possible, and the details of an efficient competitive electricity market are neither obvious nor easy to put in place. The benefits of reform may be substantial, but they require careful attention to market design. A review of the past identifies some choices on the road ahead.

INTRODUCTION

The international experience in restructuring electricity market institutions has been reflected in the many debates and experiments in the United States. The details matter, as is illustrated by examples of both success and failure. A competitive electricity market can be a vehicle for pursuing the public interest, but only if the market structure addresses the particular characteristics of the electricity system with its complex mix of essential facilities and large network externalities. Restructuring is the better term, not deregulation. Electricity is an example of the phenomenon where introducing competition leads not to less regulation, only different regulation.[2]

---

[1] William W. Hogan is the Lucius N. Littauer Professor of Public Policy and Administration, John F. Kennedy School of Government, Harvard University and a Director of the LECG, LLC. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for American National Power, Brazil Power Exchange Administrator (ASMAE), Comision Reguladora de Energia (CRE, Mexico), British National Grid Company, Calpine Corporation, Commonwealth Edison Company, Detroit Edison Company, Duquesne Light Company, Electricity Corporation of New Zealand, Enetergy, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., Mirant Corp., National Independent Energy Producers, New England Independent System Operator, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, PJM Office of Interconnection, San Diego Gas & Electric Corporation, Sempra Energy, Southern Company, Southwest Power Pool (SPP), TransÉnergie, Transpower of New Zealand, Westbrook Power, Williams Energy Group, and Wisconsin Electric Power Company. Helpful comments received from Richard Green, Scott Harvey and Larry Ruff. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web at http://www.ksg.harvard.edu/whogan).

[2] Steven K. Vogel, Freer Markets, More Rules: Regulatory Reform in Advanced Industrial Countries,

Hence, power markets are made, they don't just happen. Importantly, the rules for access to essential facilities and pricing, to provide consistent and efficient incentives, are not mere technical details that can be deferred or left alone to be discovered through the magic of the market.

The move to greater reliance on markets rests on a belief that the market participants will respond to incentives. Markets with poorly designed institutions have provided the wrong incentives, and market participants have responded. The mistakes, once made, have been costly and difficult to fix. However, the mistakes have revealed what doesn't work. The electricity market reform process in the United States and many other countries may have reached the end of the beginning. By the turn of the millennium, efforts were well underway to move from the initial reforms of regulated markets by introducing competition, to reforms of the reforms to improve the workings of partly competitive and partly regulated markets. In at least one prominent case, California, policy was turning away from market-oriented reforms.

The complete story of electricity restructuring is a complicated matter that covers transition costs and contracts, rent seeking behavior, jurisdictional disputes, market power, and much more.[3] To focus the present discussion, the effort is not to describe the full tapestry but rather to identify the threads that relate to the matter of competitive wholesale market design. This design question centers on the complications of transmission and the implications for efficient markets. This design question is important for at least two reasons. First, it makes a big difference:

> "The practice of ignoring the critical functions played by the transmission system in many discussions of deregulation almost certainly leads to incorrect conclusions about the optimal structure of an electric power system." [4]

Second, the design challenges that arise from the special nature of electricity transmission are surprising and somewhat counterintuitive. Other problems such as cost recovery, non-discrimination, and retail competition are important, but they lend themselves to more straightforward analysis and have familiar analogues in other industries. By contrast, the special nature of electricity systems leads to the need for a seeming contradiction in terms: coordination for competition. The roots of the electricity market reform policy grow into a discussion of the essential ingredients for competition and the implications for further reforms.

THE ROOTS OF ELECTRICITY RESTRUCTURING

The motivation for electricity restructuring has been slightly different in different countries. In the United Kingdom, for example, privatization of a state owned enterprise reinforced the ideology of the Thatcher government and its interest in reducing the costs of domestic coal subsidies.[5] Similar ideological and political explanations can be found from Norway to New

---

Cornell University Press, 1996, p. 3. Willis Emmons, The Evolving Bargain: Strategic Implications of Deregulation and Privatization, Harvard Business School Press, 2000, p. 6.

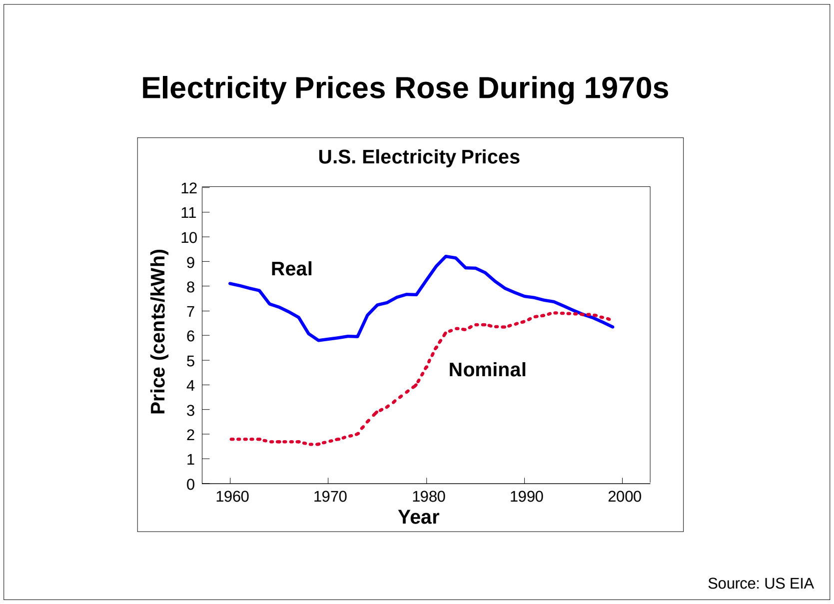
3 Paul L. Joskow, "Deregulation and Regulatory Reform in the U.S. Electric Power Sector," in Deregulation of Network Industries: The Next Steps (S. Peltzman and Clifford Winston, eds.), Brookings Press, 2000.

4 Paul L. Joskow and Richard Schmalensee, Markets for Power: An Analysis of Electric Utility Deregulation, MIT Press, 1983, p. 63.

5 Willis Emmons, The Evolving Bargain: Strategic Implications of Deregulation and Privatization, Harvard Business School Press, 2000, p. 109.

Zealand. However, there has been a common theme of growing disaffection with the electricity market model of the past and a belief or hope that the success found in "deregulation" of other industries, such as airlines or telephones, could be repeated in the case of electricity production and delivery.

In the United States, the push for restructuring electricity markets accumulated from a number of related factors. The old model, stylized as a vertically integrated monopoly with a regulated franchise, had served the country well for many years. But by the end of the 1960s, the story started to change. Until then, improved technology and further exploitation of economies of scale and scope had meant that electricity could be provided with constant or declining prices, in real and nominal terms. Meanwhile the regulated utilities enjoyed high returns and the quiet life that Hicks described as the best of all monopoly profits. All that changed in the 1970s. The oil crisis and resulting higher fuel prices, combined with higher inflation, meant that electricity prices would have to be increased to cover costs. With the apparent exhaustion of economies of scale and scope, and greater attention to environmental impacts, new investments, especially in nuclear power, were suddenly more expensive than the existing stock of generating plants. As shown in the figure, the trend in electricity prices reversed in a dramatic way, and prices were up sharply in both real and nominal terms.



The nuclear accident at Three Mile Island punctuated the transition.[6] By the beginning of the 1980s, disaffection had grown with the electric utility industry and the traditional model of the vertically integrated monopoly. The wheels were in motion for dramatic changes in the industry. The scope and surprise of the problems were captured in the massive 1983 bond default of the Washington Public Power Supply System, WPPSS, ironically know as "Whoops."[7] Previously unthinkable, these events foreshadowed other financial crises and bankruptcies in the previously stable electric utility industry.

In that same year, Joskow and Schmalensee described both the accumulating disaffection

---

[6] On March 28, 1979, the Three Mile Island Unit 2 (TMI-2) nuclear power plant near Middletown, Pennsylvania suffered a partial core melt. Nuclear Regulatory Commission, Annual Report - 1979, NUREG-0690, Washington DC.

[7] In 1983 Washington Public Power Supply System defaulted on $2.25 billion of bonds due to inability to complete five nuclear reactors. "It was the largest municipal bond default in U.S. history." David Mhyra, Whoops!/WPPSS: Washington Public Power Supply System Nuclear Plants, McFarland, 1984, p. 1-2.

with the old utility industry model and the challenges then ahead for "utility deregulation."[8] Their analysis holds up well in retrospect. While recognizing the failings of traditional regulation, Joskow and Schmalensee analyzed the difficulties of using markets given the complex technology of the electricity system. "The close physical linkages of the components of a modern power system raise serious externality problems."[9] The authors laid out a series of scenarios, intended to span the range of plausible deregulation scenarios. In the event, their most radical alternative was more conservative than the patterns that emerged in the complicated policy dance as the electricity industry moved towards greater reliance on markets and competition.

A major factor that reinforced the interest in markets grew from an initially obscure element of the Public Utilities Regulatory Policies Act of 1978 (PURPA).[10] As part of a comprehensive effort to address an "energy crisis," PURPA included many elements dealing with conservation and natural gas. Little noticed was the creation of a special class of non-utility generators who could build small power plants and co-generation facilities, known as "qualifying facilities" (QF). Section 210 of PURPA required that traditional utilities purchase electricity from QF facilities at prices set at administrative estimates of the utilities' avoided costs. For many reasons, these estimates of the avoided cost, some set in legislation such as New York's infamous "six cent law," were high enough to produce a market reaction that surprised and almost overwhelmed the regulatory process.[11] The original expectation was that QF supplies would be atypical and represent a small fraction of the market. In practice, in the most aggressive states the high administrative prices coupled with the ingenuity of new entrepreneurs who stormed into the market were enough to create a massive problem of excess capacity.

The fallout from PURPA produced may things. Regulators in states as different as California and Maine scrambled to change the rules and lower estimates of avoided costs in order to avoid the new costs they were creating. However, more importantly for the larger restructuring effort in the United States and elsewhere, the unexpected success of PURPA in stimulating new supplies put a stake through the heart of the old view that independent power producers could not provide cost-effective and reliable supplies. Furthermore, the new generating companies became effective at lobbying to put pressure on the system to relax the QF rules and, eventually, to allow independent generators to build power plants without any special restrictions. A new industry emerged.

The opportunities that could be seen in the success of the QFs created a new vision for the electricity market. The idea blossomed that a fully competitive electricity generation industry could

---

8 Paul L. Joskow and Richard Schmalensee, Markets for Power: An Analysis of Electric Utility Deregulation, MIT Press, 1983.

9 Paul L. Joskow and Richard Schmalensee, Markets for Power: An Analysis of Electric Utility Deregulation, MIT Press, 1983, p. 41

10 Public Utilities Regulatory Policies Act of 1978, 16 U.S.C. § 2601 et seq.

11 In 1981, New York law required payment of six cents per kilowatt-hour ($60/MWh) for QF power; N.Y. Pub. Serv. Law Section 66-c.1. In California, the QF "standard offer" solicitations at avoided costs were so oversubscribed that the California regulators sought coordinated procurement through the "Biennial Resource Plan Update" (BRPU) which required utilities to put their planned new generation out to bid. In the end, the regulators never approved new plant construction in the BRPU proceeding; Southern California Edison Company, et al., (1995) 70 FERC ¶ 61, 215, at p. 61,677. The collapse of the BRPU process played a prominent role in the move to reform regulation in California.

set the framework for the future. Given access to the transmission grid and other essential facilities, these generators could move their power to utilities that would buy in a competitive wholesale market. The term of art for such movement of wholesale power across the territories of multiple utilities was "wheeling." The spirit of the time supporting the introduction of competition was captured in the main title of an important study of the Congressional Office of Technology Assessment (OTA), "Electric Power Wheeling and Dealing."[12] Like with the earlier scenarios of Joskow and Schmalensee, the OTA anticipated competition limited to the wholesale purchases by existing utilities. However, the debate raged on with a great deal of attention as to whether the open competitive market would extend to final customers through "retail wheeling" or be limited to wholesale transactions. The proposal to include retail transactions was opposed by most electric utilities, who lobbied to "just-say-no" to retail wheeling.[13] The attendant debate diverted attention from more fundamental market design issues.

The reforms in the United States accelerated given the observation of the advance of electricity restructuring in England and Wales. In 1989, the British government launched the restructuring and privatization of the state owned Central Electricity Generating Board to include separation of the ownership and operation of generation, transmission and distribution.[14] The British policy also included eventual extension of competition and choice for retail customers. Similar innovations followed in Norway in 1991.[15] Chile had been the first to launch a major effort to reorganize electricity markets in 1982, but the Chilean model had more influence in Latin America than in England and the United States.[16] There were many sources of reform proposals, and the ideas were in the air.

At the next major transition in the United States, the "just-say-no" utilities appeared to win the battle. The breakthrough legislation in the Energy Policy Act of 1992 (EPAct) explicitly disavowed any extension of open competitive markets to retail customers.[17] However, EPAct included a number of other provisions that ultimately had profound effects. The law expanded the scope of QFs by creating a new class of exempt wholesale generators (EWG), essentially power producers that could be either independent or affiliated with traditional utilities, but would be spared the usual restrictions under the regulations for holding companies.[18] Furthermore, EPAct required utilities to give third parties access to their transmission systems in order to facilitate wholesale

---

12 Office of Technology Assessment, United States Congress, Electric Power Wheeling and Dealing: Technological Considerations for Increasing Competition, Washington DC, May 1989.

13 The distinctive phrase achieved its original popularity in a campaign against the use of illegal drugs.

14 The Electricity Act of 1989 set the stage for privatization and launch of the new market and the attendant electricity "Pool" in 1990. David M. Newbery, Privatization, Restructuring, and Regulation of Network Utilities, MIT Press, 1999, p. 202.

15 David M. Newbery, Privatization, Restructuring, and Regulation of Network Utilities, MIT Press, 1999, p. 246.

16 Hugh Rudnick, Ruy Varela, and William W. Hogan, "Evaluations of Alternatives for Power System Coordination and Pooling in a Competitive Environment," IEEE Transactions on Power Systems, 1996.

17 Energy Policy Act of 1992, Public Law 102-486.

18 Public Utility Holding Company Act of 1935, Public Law 74-333 (PUHCA). The law provides for regulation under the Securities and Exchange Commission, and was an earlier reform designed to restrict the activities of utility holding companies.

trading and competition. Transmission open access, largely still undefined, had become the law of the land.

Eventually, EPAct came to be seen as "…one of the most significant pieces of legislation in the history of the industry."[19] But the initial expectations of the framers of the law were more modest. At the time, non-utilities provided less than 10% of the total production volume, with the vast bulk of electricity production occurring through integrated utilities that sold to their own customers or other utilities.[20] The assumption was that this arrangement would more or less continue, allowing for a modest amount of competition at the margin.[21] The assumption was wrong. The camel's nose was in the tent, and soon the whole camel followed. The introduction of a little competition created pressure for more, and the process moved aggressively to expand the opening that had been created by the small volumes from non-utilities and the requirements of open access to the transmission wires.

The expansion beyond marginal competition flowed from two parallel streams, one in the state efforts to open up retail competition and the other in the implementation of EPAct by the Federal Energy Regulatory Commission (FERC). At the state level, the most prominent initiative was in California, where the California Public Utility Commission (CPUC) responded to a perception of a growing crisis in traditional regulation and sought an alternative in greater reliance on the new EWGs and the forces of competition. The CPUC organized an extensive and extended effort to fashion a restructured industry. The CPUC staff report, known from its cover as the "Yellow Book," concluded that California should reform its regulatory program and offered alternative strategies.[22] After further public discussion, the CPUC issued its "Blue Book" plan to substantially reorganize the structure of the industry and its regulation.[23] The subsequent implementing order by the CPUC laid out detailed prescriptions for a new market design.[24] The new design borrowed much from the experience with the reforms in England and Wales. The effect of the decisions by the CPUC was to radically alter the nature of the market in California, soon separating generation, transmission and distribution and creating new institutions to coordinate the market. Further, going well beyond the EPAct, California set in motion a plan to open its retail

---

[19] Energy Information Administration, The Changing Structure of the Electric Power Industry 2000: An Update, DOE/EIA-0562(00), Washington DC, October 2000, p. 33.

[20] Energy Information Administration, The Changing Structure of the Electric Power Industry 2000: An Update, DOE/EIA-0562(00), Washington DC, October 2000, p. 117.

[21] An observation from Congressman Philip Sharp, Chair of the Energy and Power Subcommittee of the House Committee on Energy and Commerce from 1981-95, and a principal author of EPAct.

[22] California Public Utility Commission, Decision 92-09-088, W4, 43, "Order Instituting Investigation on the Commission's Own Motion to Implement the Biennial Resource Plan Update Following the California Energy Commission's Seventh Electricity Report," September 16, 1992. California Public Utility Commission, Division of Strategic Planning (DSP), "California's Electric Services Industry: Perspectives on the Past, Strategies for the Future," February 3, 1993

[23] California Public Utility Commission, "Order Instituting Rulemaking on the Commission's Proposed Policies Governing Restructuring California's Electric Services Industry and Reforming Regulation and Order Instituting Investigation on the Commission's Proposed Policies Governing Restructuring of California's Electric Services Industry and Reforming Regulation," Docket Nos. R.94-04-031 and I.94-04-032, April 20, 1994.

[24] California Public Utility Commission, Decision 95-12-063 December 20, 1995, as amended y D.96-01-009, January 10, 1996.

markets to competition.

These innovations in California went beyond the scenarios of Joskow and Schmalensee, the scenarios of the OTA, or the expectations of the framers at the time of the passage of the EPAct. Although the story is more complicated than this cursory summary, and the results had different effects in different states, there is little doubt that the California example had a profound impact. It changed the national perception of electricity market reform from one of limited competition at the edges of the wholesale market to full blown separation of the functions of utilities into many independent pieces with unbundled supply and pricing. [25]

While California was pushing forward its radical proposals, the parallel activities of the FERC expanded the scope and conditions of what would prove to be a critical element of the evolving wholesale market, namely access to the transmission grid. Following an extensive series of paper filings and technical conferences, the FERC issued its transmission open access provisions in Order 888 with its companion information systems mandate.[26] The FERC advertised the importance of this landmark order by assigning the identifying number that coincided with the address of its new headquarters in Washington.[27] The intent in implementing the principle of open access was to give everyone equal rights to use the transmission grid. The regulatory device would be to require comparability of service. The basic structure of the industry would remain, with vertically integrated utilities, but each utility would be required to provide transmission service in a manner that was "comparable" to the transmission service it provided to itself. In effect, this would separate the transmission function from the rest of the utility. The hope was that non-discrimination would be the key to ensuring the necessary support for the competitive market.

The decisions under Order 888 coupled with the unfolding reforms in California and other states, reinforced by the examples in other countries, soon swept away the more limited scope for competition as anticipated by the framers of EPAct in 1992. By 1996 in the United States, it was clear that to some degree and in some regions, a restructured industry would include retail competition, unbundled services, and complete unpacking of the generation, transmission and distribution activities through either separation of the functions or separation of the companies. The generation and retail supply sectors would be treated as competitive industries. Distribution wires would continue under traditional monopoly franchise regulation. And somehow the essential facility in between – transmission – would be available on an open access basis.

Even if never fully achieved, this more radical scenario presented a problem for the development of electricity markets. The electric system is complex and there are many functions that must be performed but that are unlike the requirements of other industries. Most prominent is the necessity to maintain system balance of supply and demand, moment to moment, given the

---

[25] For a further elaboration of state case studies, see Energy Information Administration, The Changing Structure of the Electric Power Industry 2000: An Update, DOE/EIA-0562(00), Washington DC, October 2000, pp. 82-90.

[26] Federal Energy Regulatory Commission, Order No. 888, Docket Nos. RM95-8-000 and RM94-7-001 "Promoting Wholesale competition Through Open Access Nondiscriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities," Final Rule issued on April 24, 1996. " Open Access Same-Time Information System (formerly Real-Time Information Networks) and Standards of Conduct," FERC Order 889, Final Rule, Washington, DC, April 24, 1996.

[27] 888 First Street N.E., Washington DC, 20426. The zip code is still available for a future reform.

inability to store electricity at a reasonable cost. Second, understood by electrical engineers but unfamiliar to most others, is the requirement to manage the complex externalities associated with the flow of power across constrained transmission systems. It has long been recognized that some "…pooling and coordination entity will have to be created to serve as an intermediary (both physical and financial) between individual producers of electric power and wholesale consumers (primarily distribution companies)."[28] "With large amounts of competitive or unbundled generation … explicit arrangements for coordinated dispatch and scheduling will be required."[29] As unbundling proceeded and competition expanded, the need for some entity performing these functions became ever more obvious.

This necessity was not lost on the FERC. As is its custom, the FERC included in Order 888 a review of the public comments and problem diagnoses. A close reading finds an extensive discussion of the obstacles to electricity markets created by the need for instantaneous balancing and managing the externalities of transmission usage. In particular, the FERC recognized that the traditional power "wheeling" model was built on the fiction of the "contract path."[30] In other words, the trading arrangements were based on the assumption that the power could be directed to follow a particular path in the network, in contravention of the accepted physical reality that the power would flow over every parallel path. In the old world of vertically integrated utilities, with small volumes of non-utility production, the contract path was a workable fiction for commercial purposes, and the engineers could deal separately with the physical reality. But in the new unbundled world with a growing volume of third party transactions, the traditional wheeling model would break down. A simple reading of Order 888 shows plainly that the FERC knew all this, but in the end FERC embraced the wheeling model for the expedient reason that it could not reach agreement on an alternative approach for coordinating transmission service.

The evidence of the immediate seriousness of the problems thus created was readily at hand. For example, shortly after the adoption of Order 888, the North American Electric Reliability Council (NERC), the organization responsible for system reliability, recognized that contract-path scheduling created incentives to overload the electric network system. The NERC immediately adopted transmission loading relief protocols to undo the damage whenever the system became constrained.[31] In essence, NERC created an administrative un-scheduling system to counteract the effects of the FERC-mandated scheduling system.[32] The NERC system did not work well.[33] However, something was necessary in order to keep the lights on.

---

[28] Paul L. Joskow and Richard Schmalensee, Markets for Power: An Analysis of Electric Utility Deregulation, MIT Press, 1983, p. 114.

[29] Office of Technology Assessment, United States Congress, Electric Power Wheeling and Dealing: Technological Considerations for Increasing Competition, Washington DC, May 1989, p. 133.

[30] The contract paths are redefined as "posted paths" in Federal Energy Regulatory Commission, "Open Access Same-Time Information System (formerly Real-Time Information Networks) and Standards of Conduct," Order No. 889, Final Rule, Washington, DC, April 24, 1996, p. 66.

[31] Rajesh Rajaraman and Fernando L. Alvarado, "Inefficiencies of NERC's Transmission Loading Relief Procedures," Electricity Journal, October 1998, pp. 47-54.

[32] Michael Cadwalader, Scott Harvey, William Hogan, and Susan Pope, "Market Coordination of Transmission Loading Relief Across Multiple Regions," Center for Business and Government, Harvard University, December 1, 1998.

[33] Congestion Management Working Group of the NERC Market Interface Committee, "Comparison of

Perhaps the most striking evidence that the problems created by the contract-path approach were both serious and immediate appeared in the words of the FERC itself. Remarkably, on the very day the FERC issued its landmark open access tariff in Order 888, the FERC issued a companion notice of proposed rulemaking for a new transmission capacity reservation tariff (CRT). The notice included the stunning preamble:

> "The proposed capacity reservation open access transmission tariff, if adopted, would replace the open access transmission tariff required by the Commission ..."[34]

Apparently after years of deliberation and mountains of paper, the FERC knew that what it had just wrought would fail, and something else would be required. The proposed capacity reservation tariff would create completely new arrangements for coordinated dispatch and scheduling.[35] The CRT proposal received a generally negative review from the industry. It soon disappeared from the FERC agenda. However, as we shall see below, the CRT later reappeared in another guise.

The discussion of the CRT was overtaken by events. Most prominent was the parallel rise of the independent system operator (ISO). The prescient predictions of the earlier analyses of electricity market restructuring were born out by the arrival of new institutions for coordinating system dispatch and use of the transmission grid. Given current electric technology, some such entity is necessary. The question is not whether there should be a system operator, the only meaningful question is what should be rules and protocols that the system operator should follow in support of a competitive market.

The market structures of England and Wales, Norway and many other countries depend on coordination through system operators.[36] Following their lead and encouraged by the FERC, new ISOs appeared in California (CAISO), the Pennsylvania-New Jersey-Maryland Interconnection (PJM), New York (NYISO), New England (ISONE), Texas (Electric Reliability Council of Texas-ERCOT), and the Midwest (MISO). More were in the planning stages. The recognition of the limitations of the landmark Order 888, the evident necessity of such coordinating organizations, and the pressing requirements for specifying the rules, transformed slowly into another extended set of hearings and filings at the FERC under the rubric of a new name, the Regional Transmission Organization (RTO). The FERC was asking: what were these ISOs, or RTOs, or Poolcos, or similar entities under a myriad of new names, supposed to do exactly? The FERC then issued a new order with another signature numbering, Order 2000, referred to here as the Millennium Order.[37] This order was the CRT born again in greatly expanded form. To develop the implications of these reforms, it is necessary to consider further the requirements for supporting a competitive electricity

---

System Redispatch Methods for Congestion Management," September 1999.

[34] Federal Energy Regulatory Commission, "Capacity Reservation Open Access Transmission Tariffs," Notice of Proposed Rulemaking, RM96-11-000, Washington DC, April 24, 1996, p. 1.

[35] Scott M. Harvey, William W. Hogan, and Susan L. Pope, “Transmission Capacity Reservations and Transmission Congestion Contracts,” Center for Business and Government, Harvard University, June 6, 1996, (Revised March 8, 1997).

[36] For a thorough discussion of the early designs and roles of contracting, see Sally Hunt and Graham Shuttleworth, Competition and Choice in Electricity, John Wiley and Sons, 1996.

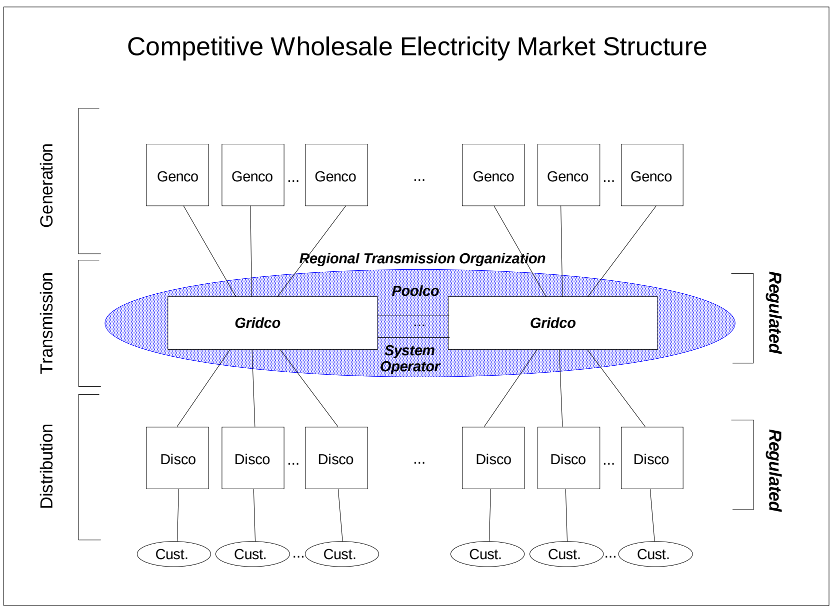
[37] Federal Energy Regulatory Commission, "Regional Transmission Organizations," Order No. 2000, Docket No. RM99-2-000, Washington DC, December 20, 1999.

market.

THE ESSENTIAL MARKET INGREDIENTS

Real markets are complicated by imperfections with information asymmetries, transaction costs and market power. The best we can hope for is workable competition. However, even if we assume that buyers and sellers in the wholesale electricity generation market act as pure price takers, the competitive case, the task of market design confronts special difficulties in the circumstances of electricity markets.

The central problem in the development of competitive electricity markets arises from the need for a system operator who can manage the complex short-term interactions in the network and maintain system reliability.[38] There must be a system operator. The only open questions are about the rules the system operator will apply and the governance of its activities. Given that the subject is internalizing externalities, there are winners and losers. Given that the subject is complex, it confronts conflicting ideologies. The topic has proven to be highly controversial. Nevertheless, the development of Independent System Operators has proceeded steadily in the worldwide restructuring of electricity markets. There are significant advantages in this approach. Control of the use of the transmission grid means control of the dispatch, at least at the margin, because adjusting the dispatch is the principal (or, in some cases, only) means of affecting the flow of power on the grid. That this system operator should also be independent of the existing electric utilities and other market participants is attractive in its simplicity in achieving equal treatment of all market participants. The ISO provides an essential service, but does not compete in the energy market.



The process of restructuring wholesale electricity markets in the United States has added to the extensive worldwide debate about the range of possible and preferred alternatives for organizing regional electricity markets. Most importantly, the FERC addressed a wide range of issues in its analysis of and orders for the design of Regional Transmission Organizations. The Millennium RTO Order covers a great deal in fashioning well-designed market institutions to serve the public

---

38 For expanded version of the argument here, see William W. Hogan, "Regional Transmission Organizations: Millennium Order on Designing Market Institutions for Electric Network Systems," Center for Business and Government, Harvard University, May 2000.

interest.

Surprisingly for an industry as capital intensive as electricity production and distribution, the essential elements are found in a consistent organization of short-run operations and the associated pricing. Difficult or otherwise intractable problems that arise in electricity markets, in both the long run and the short run, disappear or are simplified when the pieces fit together for efficient short-term operations in the context of flexible choices for market participants.

In the short run, there are critical functions that must be performed by someone. The complex network interactions in an electric grid require that there be an entity that can provide certain critical coordinating services.[39] But the implications that follow from this fact are so contentious that the discussion often becomes confused and the language strained. Here we focus on the activities of this entity as the system operator, no matter what final name we may give it.

The most obvious example of the essential services is in energy balancing. The electric system must maintain continuous aggregate balance of production and consumption. This same balance of inputs and outputs must be coordinated in a way that respects the many limits in the transmission system. Hence, not only must the aggregate inputs and outputs conform to the electrical laws that govern the interconnected grid, but the locational pattern of power production and use must honor these same laws in order to manage the flow of power within the limits of the transmission system.[40] Simultaneously, in order to maintain reliability within the security limits of the grid, various ancillary services such as spinning reserve and reactive support need close coordination and monitoring.

This coordination function is not optional. It appears in every electric system. It must be provided. And the services must be integrated with each other. The needs for reactive power and spinning reserve depend importantly on the overall pattern of power production and use. Individual market participants can produce individual elements of these services, but the fundamental coordination function requires a single entity. This is the responsibility of the system operator. And there is always a system operator.

Since the functions of the system operator are not optional, the only open question for market design is how these functions will be performed. The system operator could do a good job, meaning operating efficiently to support a competitive market. Or the system operator could do a bad job, providing the services in a way that increases costs and undermines the competitive market. The central effect of policy should be to require good design for the functions of the system operator.

A central problem appears in designing the design process. Experience indicates that reliance on voluntary agreements among market participants is not likely to be successful. Some problems, like dividing the pie, are largely political and voluntary agreement would be natural. But other problems, like designing bridges, dictate a need for careful consideration of how the pieces fit together and what is in the public interest. Electricity market design is more like the latter than the former. The Millennium Order responds to the need for coherent design that recognizes the complexity of electricity markets.

---

[39] RTO Order, p. 270. See also, William W. Hogan, "Independent System Operator: Pricing and Flexibility in a Competitive Electricity Market," Center for Business and Government, Harvard University, February 1998.

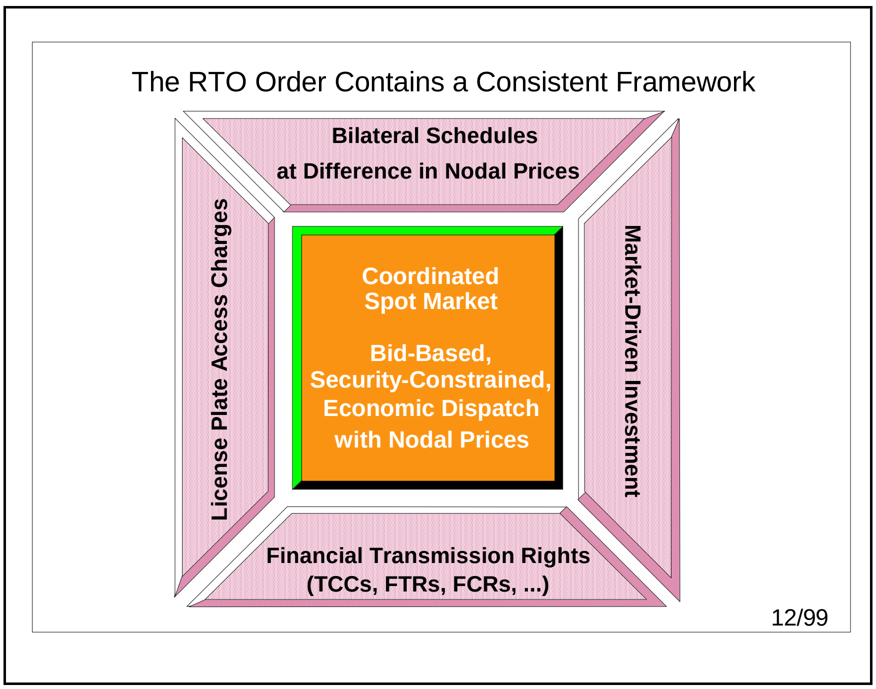
[40] RTO Order, pp. 423-424.

The example of energy balancing illustrates the point. Energy balancing and congestion management are inextricably intertwined. The best approach is to run the balancing and congestion management market as a bid-based, security-constrained economic dispatch with voluntary participation by generators and loads. The corresponding prices would be consistent with the competitive outcome and would reflect the marginal cost of meeting load at each location.

To do anything else would be to decide on providing the essential coordination services in a way that would be inconsistent with the fundamental goals of electricity restructuring and inconsistent with the basic principle of designing market institutions to support the public interest. As a matter of good public policy, we should not have an interest in market designs that raise costs and decrease the real flexibility of market participants.[41]

These same essential ingredients would provide many other benefits. Bilateral transmission schedules of great flexibility and market-responsiveness could be accommodated with the transmission usage price set consistently at the difference in the locational energy prices. There would be no bias between bilateral schedules and the coordinated spot market. The market for ancillary service acquisition and pricing could be integrated simultaneously in the economic dispatch.

From the perspective of design of institutions, the most important theme running through the Millennium Order's discussion of these characteristics and functions is the prominence of markets as the means for achieving the many goals of electricity restructuring. The key element is in the recognition of the importance of a coordinated spot market. In the Millennium Order this appears principally in the discussion of the balancing market. In particular, "[r]eal-time balancing is usually achieved through the direct control of select generators (and, in some cases, loads) who increase or decrease their output (or consumption in the case of loads) in response to instructions from the system operator."[42] To be consistent with the competitive market, it is essential that this be through a bid-based, security-constrained economic dispatch: "Proposals should ... ensure that (1) the



41 Larry Ruff, "Competitive Electricity Markets: One Size Should Fit All," The Electricity Journal, November 1999, pp. 20-35.

42 RTO Order, p. 635.

generators that are dispatched in the presence of transmission constraints must be those that can serve system loads at least cost, and (2) limited transmission capacity should be used by market participants that value that use most highly."[43]

Further, the FERC requires that everyone be able to participate in this coordinated spot market, at the efficient, and necessarily locational or nodal, prices: "The market mechanisms must accommodate broad participation by all market participants, and must provide all transmission customers with efficient price signals regarding the consequences of their transmission usage decisions."[44] In addition, "[t]he Regional Transmission Organization must ensure that its transmission customers have access to a real-time balancing market. The Regional Transmission Organization must either develop and operate this market itself or ensure that this task is performed by another entity that is not affiliated with any market participant."[45]

Efficient Pricing

Efficient pricing is a central feature of a competitive electricity market. It is essential if the benefits of a competitive market are to flow through to customers and other market participants. Pricing that is inefficient, on the other hand, will fail to signal and encourage appropriate levels of consumption and supply or the appropriate levels and locations of new generation and transmission investment.

The standard determinant of competitive market pricing is system marginal cost. This is the simple definition of the market-clearing price where supply equals demand. This production level just balances the marginal benefit of additional consumption with the marginal cost of production. Under the usual competitive assumptions, this textbook market equilibrium condition also provides the welfare maximizing economic outcome, which is the definition of economic efficiency.

The basic textbook model extends to the definition of competitive equilibrium for products across multiple locations. The same criterion applies in finding the economic, or least-cost, dispatch of the power grid given the benefits of consumption or the costs of production at each location.[46] Using the bids as the representation of these benefits and costs, the corresponding economic dispatch produces the same outcome as a competitive equilibrium. The economic dispatch accounts for system congestion and transmission losses, and thus inherently produces prices that can vary at each location by the combined effect of generation, losses and congestion. These locational prices provide proper signals for the quantity and location of new investment.

As a matter of principle, these locational prices are simply the market-clearing prices based on all the bids and the details of the requirements of network operations. Furthermore, for any given economic dispatch, it is an easy matter to determine these prices based on the bids and

---

43 RTO Order, pp. 332-333. See also p. 382.

44 RTO Order, p. 332. See also p. 743.

45 RTO Order, p. 423. See also p. 715.

46 F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R.E. Bohn, Spot Pricing of Electricity, Kluwer Academic Publishers, Norwell, MA, 1988.

the system conditions. These locational prices are in use today as an integral part of the market design in many regions.

In addition to defining the market-clearing price at each location, these locational prices provide an immediate and simple answer to the otherwise intractable question as to the appropriate marginal cost or market-clearing price of transmission use. The electric network is complicated, with the power flow dictated by the laws of physics and many system constraints. Tracing the details of transmission flow has proven to be a blind alley that has frustrated attempts to define workable methods of transmission pricing.[47] But the locational pricing approach that accompanies the coordinated spot market provides an immediate simplification of this difficult problem. In particular, transmission of a megawatt between two locations is physically equivalent to sale at the source and purchase at the destination. In equilibrium, therefore, the market-clearing price determined by the marginal cost of transmission must be the same as the net price for the combined purchase and sale transaction. In other words, the price of transmission between two locations must be just the difference in the locational prices of energy.

Since these pricing conditions are derived from first principles for a competitive equilibrium, any efficient mechanism must produce the same pricing result. It follows, therefore, that the market design requirement for a system operator with a balancing and congestion management system provides an easy solution for the efficient support of a competitive market. Economic dispatch with its locational prices defines the efficient outcome.

This is not a new idea. "Spot pricing (or real-time pricing) is another approach that has been considered for coordinating the output of generators to follow loads. … However, a lack of experience with spot pricing leaves significant uncertainties about its practical application."[48] What is new is the practical experience obtained in many countries that shows such efficient pricing to be at least practical, and perhaps essential.

When the system is constrained, the spot prices create congestion rents reflecting the opportunity cost of the constraints. Similarly, rents arise in pricing transmission losses. An issue then arises as to the best use of these transmission rentals. The basic logic is that the payment should be divorced from the marginal usage decisions, in order to preserve the incentives of efficient pricing. Further, it is intuitive that the proper recipients of the rentals should be those who are paying the transmission access charges to cover the fixed costs of the grid. This logic is consistent as far as it goes. However, as we shall see, a superior use of these rentals is in funding long-term transmission rights.

---

[47] William W. Hogan, "Flowgate Rights and Wrongs," Center for Business and Government, Harvard University, August 2000. Larry E. Ruff, "Flowgates, Contingency-Constrained Dispatch, and Transmission Rights," Electricity Journal, Vol. 14, No. 1, January/February 2001, pp. 34-55.

[48] Office of Technology Assessment, United States Congress, Electric Power Wheeling and Dealing: Technological Considerations for Increasing Competition, Washington DC, May 1989, p. 133.

Long-Term Transmission Rights

With changing supply and demand conditions, generators and customers will see fluctuations in short-run prices. Changing flows will produce changes in losses. When demand is high, more expensive generation will be employed, raising the equilibrium market prices. When transmission constraints bind, congestion costs will change prices at different locations.

Even without transmission congestion constraints or significant changes in losses, the spot market price can be volatile. This volatility in prices presents its own risks for both generators and customers, and there will be a natural interest in long-term mechanisms to mitigate or share this risk. The choice in a market is for long-term contracts.

Traditionally, and as is seen in many other markets, the notion of a long-term contract carries with it the assumption that customers and generators can make an agreement to trade a certain amount of power at a certain price. The implicit assumption is that a specific generator will run to satisfy the demand of a specific customer. To the extent that the customer's needs change, the customer might sell the contract in a secondary market, and so too for the generator. Efficient trading in a secondary market would guarantee equilibrium and everyone would face the true opportunity cost at the margin.

However, this notion of specific performance stands at odds with the operation of the short-run market for electricity. To achieve an efficient economic dispatch in the short-run, the dispatcher must have freedom in responding to the bids to decide which plants run and which are idle, independent of the provisions of long-term contracts. And with the complex network interactions, it is impossible to identify which generator is serving which customer. All generation is providing power into the grid, and all customers are taking power out of the grid. It is not even in the interest of the generators or the customers to restrict their dispatch and forego the benefits of the most economic use of the available generation. The short-term dispatch decisions by the system operator are made independent of and without any recognition of any long-term contracts. In this way, where the disconnect between operations and contracts is not only feasible but necessary, electricity is not like other commodities.

This dictate of the physical laws governing power flow on the transmission grid does not preclude long-term contracts, but it does change the essential character of the contracts. Rather than controlling the dispatch and the short-run market, long-term contracts focus on the problem of price volatility and provide a price hedge not by managing the flow of power but by managing the flow of money. The short-run prices provide the right incentives for generation and consumption, but create a need to hedge the price changes. Recognizing the operation of the short-run market, there is an economic equivalent of the long-run contract for power that does not require any specific plant to run for any specific customer.

Consider the case first of no transmission congestion and no losses. In this circumstance, it is possible to treat all production and consumption as at the same location. Here the natural arrangement is to contract for differences against the equilibrium price in the market. A customer and a generator agree on an average price for a fixed quantity, say 100 MW at $50 per megawatt-hour (MWh). On the hour, if the spot price is $60, the customer buys power from the spot market at $60 and the generators sells power for $60. Under the contract, the generator owes the customer $10 for each of the 100 MW over the hour. In the reverse case, with the spot price at $30, the customer pays $30 to the system operator, which in turn pays $30 to the generator, but now the customer owes the generator $20 for each of the 100 MW over the hour.

This then is the familiar "contract for differences (CFD)."[49] It is a forward contract like those found for other traded commodities but discovered anew for electricity as an innovation in the market in England and Wales. The CFD allows for long-term contracts without direct contract administration by the system operator.

In effect, the generator and the customer have a long-term contract for 100 MW at $50. The contract requires no direct interaction with system operator other than for the continuing short-run market transactions. But through the interaction with system operator and the coordinated spot market, the situation is even better than with a long-run contract between a specific generator and a specific customer. For now if the customer demand is above or below 100 MW, there is a ready and an automatic spot market where extra power is purchased or sold at the spot price. Similarly for the generator, there is an automatic spot market for surplus power or backup supplies without the cost and problems of a large number of repeated short-run bilateral negotiations with other generators. And if the customer really consumes 100 MW, and the generator really produces the 100 MW, the economics guarantee that the average price is still $50. Furthermore, with the contract fixed at 100 MW, rather than the amount actually produced or consumed, the long-run average price can be guaranteed without disturbing any of the short-run incentives at the margin. Hence the long-run contract is compatible with the short-run market.

In the presence of transmission congestion and losses, the generation contract is necessary but not sufficient to provide the necessary long-term price hedge. A bilateral arrangement between a customer and a generator can capture the effect of aggregate movements in the market, when the single market price is up or the single market price is down. However, transmission congestion and losses can produce significant movements in price that are different depending on location. If the customer is located far from the generator, transmission congestion might confront the customer with a high locational price and leave the generator with a low locational price. Now the generator alone cannot provide the natural back-to-back hedge on fluctuations of the short-run market price. Something more is needed.

Transmission congestion and losses in the short-run market raise another related and significant matter for the system operator. For example, in the presence of congestion, revenues collected from customers will substantially exceed the payments to generators. The difference is the congestion rent that accrues because of constraints in the transmission grid. At a minimum, this congestion rent revenue itself will be a highly volatile source of payment to the system operator. At worse, if the system operator keeps the congestion revenue, incentives arise to manipulate dispatch and prevent grid expansion in order to generate even greater congestion rentals. System operation is a natural monopoly and the operator could distort both dispatch and expansion. The same would apply to rents on losses. If the system operator retains the benefits from transmission rentals, this incentive would work contrary to the goal of an efficient, competitive electricity market.

The convenient solution to both problems – providing a price hedge against locational congestion differentials and removing the adverse incentive for the system operator – is to re-distribute the congestion and net loss revenue through a system of long-run financial

---

[49] Sally Hunt and Graham Shuttleworth, Competition and Choice in Electricity, John Wiley and Sons, 1996, pp. 119-132.

transmission rights (FTR) operating in parallel with the long-run generation contracts.[50] Just as with generation, it is not possible to operate an efficient short-run market that includes transmission of specific power to specific customers. However, just as with generation, it is possible to arrange an FTR that provides compensation for differences in prices, in this case for differences in the congestion and marginal loss costs between different locations across the network.

The FTR for compensation would exist for a particular quantity between two locations. The generator in the example above might obtain an FTR for 100 MW between the generator's location and the customer's location. The right provided by the contract would not be for specific movement of power but rather for payment of the price difference. Hence, if a transmission constraint caused the price to rise to $60 at the customer's location, but remain at $50 at the generator's location, the $10 difference would be the congestion rental. The customer would pay the $60 for the power. The settlement system would in turn pay the generator $50 for the power supplied in the short-run market. As the holder of the FTR, the generator would receive $10 for each of the 100 MW covered under the FTR. This revenue would allow the generator to pay the difference under the generation contract so that the net cost to the customer is $50 as agreed in the bilateral CFD power contract. Without the FTR, the generator would have no revenue to compensate the customer for the difference in the prices at their two locations. The FTR completes the package.

As with the familiar generation contract-for-differences, the FTR leaves undisturbed the marginal incentives for efficient operations. The FTR is defined for a fixed quantity. If actual usage exactly matches this quantity, the FTR provides a perfect transmission price hedge. But if usage exceeds this FTR quantity, there is no hedge for the incremental volume and the full incentive effect of efficient pricing applies. Likewise, if usage should be below the FTR volume, the payment would apply to the full FTR quantity, so the owner would see the proper marginal incentive to reduce transmission use.

These FTRs are equivalent to perfectly tradable physical transmission rights in a system that has parallel flows. Parallel connections increase system reliability, but create otherwise difficult problems in defining and using transmission rights.

If a simple feasibility test is imposed on the FTRs awarded to customers, the aggregate congestion payments received through the spot market will fund the payment obligations under the FTRs. Still, the transmission prices paid and received will be highly variable and load dependent. Only the system operator will have the necessary information to determine these changing prices, but the information will be readily available embedded in all the spot market locational prices. The FTRs define payment obligations that guarantee protection from changes in the transmission rentals.[51]

Given the availability of this coordinated spot market and these efficient locational prices,

---

[50] All current implementations of FTRs use congestion but not loss rentals. They are also known as Transmission Congestion Contracts (TCC), as in New York, and Financial Congestion Rights (FCR), in New England.

[51] Scott M. Harvey, William W. Hogan, and Susan L. Pope, "Transmission Capacity Reservations and Transmission Congestion Contracts," Center for Business and Government, Harvard University, June 6, 1996, (Revised March 8, 1997).

market participants could schedule bilateral transactions or rely on trade through the spot market. The differences in locational prices would define the opportunity costs of transmission, giving rise to the creation of financial transmission rights.[52] Payment for the existing grid would appear in part as access charges, including the use of the "license plate" approach with region-specific access charges.[53]

These are the most important elements. These define the functions of the essential system operator. These are not mere technical details, and they have far-reaching implications for how, and how well, the market works. The rules for access to the limited capacity of the transmission system stand at the core of all other issues.

Market Power

The problem of market power remains as an important policy issue in electricity markets. The discussion here about the essential ingredients for market design addresses the institutions needed to support a competitive generation market. The full treatment of market power is a complex issue. Furthermore, the argument would be that market design is not the best tool for mitigating market power. An examination of the implications of market power would take us too far afield. For the present purposes, recognize that the coordinated spot market design itself will not eliminate market power. Substantial market power would call into question any proposal to rely on markets for generation.

If there is significant exercise of traditional market power through withholding of generation, this has important policy implications. The preferred response would be bid caps targeted at those exercising market power in the short-run and divestiture in the long-run, and this action alone might be sufficient to moderate the average price impacts. However, if the explanation for market problems lies elsewhere, the policy implications would be different. If scarcity and higher costs are the dominant forces, bid caps on large suppliers and divestiture would have little, maybe no, impact on the outcome of prices and production. Most importantly, price caps that appear more justifiable in the presence of traditional market power become exactly the wrong approach in dealing with scarcity.[54]

REFORMS OF REFORMS

The theory of the case for the market design with an efficient coordinated spot market run by the system operator is by now well supported by practical experience. The main ingredients of the coordinated spot market with locational pricing exist in many parts of the world as diverse as Chile and New Zealand, and the combined package with FTRs has been operating successfully in PJM since 1998. The same general design has been adopted in New York,[55] and embraced as a

---

52 RTO Order, pp. 382-383.

53 RTO Order, p. 524.

54 For a further discussion of market power issues, see Scott M. Harvey and William W. Hogan, "On The Exercise Of Market Power Through Strategic Withholding In California," Center for Business and Government, Harvard University, April 24, 2001.

55 New York began operation under this market design in November 1999.

reform in New England.

Efficient pricing, in particular, is especially important in markets that allow participant choices. Almost by definition, any approach other than economic dispatch with nodal pricing will produce prices that are not consistent with market equilibrium. Inevitably, this inconsistency will drive the monopoly system operator or grid owner to intervene in the market. The problems that arise when we do anything else are apparent in various experiments where supposed simplifications produced predictable problems. The successes did not come immediately or easily, and success is not found everywhere. There have already been reforms of reforms, and more will follow. However, the outcome is uncertain. The delay in implementing good RTO designs throughout the United States leaves the restructuring process vulnerable. Other countries are still struggling with the core issues. Review of a few prominent cases that go beyond the initial reforms illustrates the general argument.

PJM Interconnection

The debate over transmission usage and transmission pricing in PJM provides a stark illustration of the difficulty and the challenge of market design.[56] In March of 1997, the FERC approved an interim transmission access and pricing system to operate in conjunction with a real-time spot market coordinated through the PJM ISO. Faced with opposition to a full locational pricing and congestion charging mechanism for actual use of the system, the FERC endorsed the locational approach in principle but adopted temporarily an alternative model based on a single market clearing price (MCP). The MCP approach minimized the importance of transmission congestion and rejected the locational pricing model as too complicated and unnecessary. Instead, the MCP model would treat the entire PJM system as a single zone.

In essence, much like the approach in England and Wales, the MCP model priced all transactions through the spot-market at the "unconstrained" price, based on a hypothetical dispatch. To the extent that the actual dispatch encountered transmission constraints, the MCP model would pay the more expensive generators to run and average these congestion costs over all users.

The model included two other notable features. First, in the face of transmission congestion, the generators that were constrained not to run would be paid nothing, even though they had bids below the "unconstrained" price. Unlike in the case of England and Wales, there was objection to adopting any system that depended on paying generators not to generate power, with the attendant discrimination and perverse incentive effects. Second, market participants had the option to schedule bilateral transactions separately from the bid-based economic dispatch of the ISO, with a separate payment for their share of the total congestion cost. This flexibility to use bilateral transactions or to participate in the coordinated spot market was a major design objective not to be abandoned.

This pricing system is representative of a zonal approach, and has much in common with zonal systems adopted elsewhere in the world.[57] However, should the system become

[56] For details, see William W. Hogan, "FERC Policy on Independent System Operators: Supplemental Comments," Federal Energy Regulatory Commission, Docket No. PL98-5-000, Washington DC, May 1, 1998.

[57] Here the issue is pricing for transmission congestion. The recovery of embedded costs of transmission

constrained, the two exceptional features noted above would create a powerful and perverse incentive. If there were no transmission constraints, there would be no transmission congestion and everything would work as with the locational pricing system. But when congestion appeared, everything would be different. The supporters of the zonal approach argued that the total cost of congestion would be small, summed over the year, and therefore any inefficiencies could be safely ignored.

Ignoring a difference between prices and marginal costs is a safe practice in a regulated world without flexibility and choice. The incentives don't matter and the small costs get lost in the larger system. Inconsistent pricing can work inside the closed black box of the vertically integrated system. But the cost of ignoring a gap between prices and marginal costs in the world of choice can be large indeed. Witness the events when the PJM system became constrained, starting in June of 1997.

The data for a representative constrained dispatch found the marginal cost in eastern PJM at about $89 per MWh, when at the same time the marginal cost in the west was $12 per MWh and the "unconstrained" price was approximately $29 per MWh. The incentives were clear. A customer could buy from the spot-market dispatch at $29, or it could arrange a bilateral transaction with a constrained-off generator in the west at a price closer to $12.[58] The small average congestion cost would be the same either way, and would not affect the choice. The choice, therefore, presented a test for generators.

Faced with these incentives, constrained-off generators passed the test. They quickly arranged bilateral transactions and scheduled their power for delivery, thereby exceeding the transmission limits. This, in turn, required the ISO to constrain the output from some other generator, who would then follow the same direct path to a bilateral schedule rather than sit idle and collect nothing. Soon the ISO had no more controllable generating units with which to manage the transmission constraints. Unable to fix the perverse pricing incentives, the ISO resorted to administrative mechanisms to prohibit bilateral transactions or declare a "minimum" generation emergency during the peak generation period. In effect, while restructuring to facilitate a market, the unintended consequences of superficially simple pricing spawned administrative rules to prohibit the market from responding to the price incentives when they mattered most. Shackled with inconsistent pricing rules, the ISO had to resort to direct preemption of market choices.

The point was made in a dramatic way. The important issue is not the total cost of congestion, which may be small on average. The point is the incentives at the margin when the system is constrained. In designing the rules for transmission pricing and access for a competitive market, it matters little what the rules are for periods when the system is unconstrained. The important question is how the rules deal with the market when the system is constrained. Even if the total cost of congestion might be modest over the year, the gap between $29 and $12, or $89 and $12, is more than sufficient to get the attention of market participants. Given the margins in this business, they will change their behavior for $1. And the changes in behavior can substantially affect system operations; in fact, the whole point of electricity

---

investment through access charges is a separate matter that is amenable to zonal approach.

[58] Power Markets Week, September 1, 1997, p. 13.

restructuring is that changes in behavior can affect system operations and lead to different patterns of electricity use and investment.

In the locational pricing system, the perverse incentives would not arise. Given the same facts as above, the locational prices would equal the marginal costs. Those customers purchasing power from the spot market in the east would have seen $89 as the price. True, they could have arranged a bilateral transaction with a generator in the west, paying $12 for the energy. But they would then face a transmission charge of $77 ($89-$12), making them indifferent at the margin, just as intended. Likewise, customers in the west would pay $12 and have no incentive to change. Every generator would be producing at its short-run profit maximizing output, given the prices. The market equilibrium would support the necessary dispatch in the presence of the transmission constraints. Spot-market transactions and bilateral schedules would be compatible. Flexibility would be allowed and reliability maintained consistent with the choices of the market participants.

The PJM ISO was fully aware of the perverse incentives of zonal congestion pricing and the problems they created, but without the authority to change the pricing rules it had no alternative but to restrict the market. Faced with this reality, the FERC acted to approve the locational pricing system that became operational in PJM at the beginning of April of 1998. The developing experience should be better understood to avoid the pitfalls of the complicated zonal "simplification." The subsequent successful experience in PJM has demonstrated the practical importance of locational pricing.[59] The PJM ISO determines locational prices for over two thousand locations every five minutes. Trading hubs are included and the western hub has become a major market center. FTR auctions occur ever month. Congestion is common. generators are building where generation is valuable. The PJM system works with both a real-time and a day-ahead market.[60]

New England

There are many ways that things can go wrong. The PJM 1997 experiment with a zonal pricing system collapsed as soon as the system became constrained. Subsequently, New England adopted a similar MCP approach but without the flexibility for participants to self-schedule to counteract dispatch instructions. However, New England found that the one-zone congestion pricing system created inefficient incentives for locating new generation.[61] Faced with uniform pricing, generators preferred to build where costs were low rather than where value was high. To counter these price incentives, New England proposed a number of limitations and conditions on new generation construction. Following the FERC's rejection of the resulting barriers to entry for new generation in New England, there developed a debate over the preferred model for managing

---

59 Two of the original sponsors of the MCP plan, Philadelphia Electric Company and Enron, subsequently became active supporters of the PJM locational pricing market model.

60 PJM Interconnection. L.L.C. For further details on the experience in PJM, see William W. Hogan, "GETTING THE PRICES RIGHT IN PJM. Analysis and Summary: April 1998 through March 1999, The First Anniversary of Full Locational Pricing," April 2, 1999, available through the author's web page; and the earlier discussion in the Electricity Journal, September 1998, pp. 61-67.

61 New England Power Pool, 85 FERC Para 61,141 (1998). For a critique of the previously proposed one-zone congestion pricing system, see Peter Cramton and Robert Wilson, "A Review of ISO New England's Proposed Market Rules," Market Design, Inc., September 9, 1998.

and pricing transmission congestion.[62] One zone was not enough, but perhaps a few would do?

The extended conversation amounted to a complete replay of all the market design issues, going well beyond the issue of congestion alone. In the end, New England proposed to go all the way to a locational pricing system. The revised model included a new coordinated spot market, locational pricing, and financial transmission rights.[63] Hence, the ISONE reforms of reforms would produce a market design that is similar to that operating in New York and PJM. The three ISOs then joined with the Ontario market operator in a memorandum of understanding to coordinate the operation of their markets and resolve seams issues.[64] Subsequently, ISONE and PJM announced an agreement for ISONE to adopt the PJM market design, protocols, and certain software.[65]

New Zealand

In many ways, the New Zealand market design has been at the forefront of best practice. Furthermore, the electricity reform process in New Zealand involved extensive consideration of the essential ingredients of market design and the experience in other countries. The New Zealand electricity market provides fundamental design elements needed to support competition in generation and supply. A key feature of any such market is the use of a coordinated spot market to handle balancing, transmission usage and security requirements. The New Zealand spot market includes a bid-based, security-constrained, economic dispatch with fully locational prices for real-time decisions. The bids summarize the preferences of the market participants and ensure that the final dispatch choices respect those preferences. The security constraints preserve the conditions needed to ensure reliable operations. The principles of economic dispatch define both the traditional engineering practice and the results of a competitive equilibrium. In this regard, the New Zealand model for real-time operations is aligned with the best international practice for a competitive electricity market.[66]

Nevertheless, motivated in large part with concerns over the results of retail competition, New Zealand has reconsidered its reforms and revisited the issues of electricity market design.[67] There are special issues in New Zealand, particularly its distinctive attempt to create regulation

---

62 Federal Energy Regulatory Commission, New England Power Pool Ruling, Docket No. ER98-3853-000, October 29, 1998.

63 ISO New England, "Congestion Management System and a Multi-Settlement System for the New England Power Pool," FERC Docket EL00-62-000, ER00-2052-000, Washington DC, March 31, 2000. The proposal includes full nodal pricing for generation and, for a transition period, zonal aggregation for loads. Federal Energy Regulatory Commission, “Order Conditionally Approving Congestion Management and Multi-Settlement Systems,” Docket No. EL00-62-000, June 28, 2000.

64 PJM, NYISO, ISONE, IMO Press Release, "Ontario’s IMO and U.S. Independent System Operators Sign Agreement To Coordinate Inter-Regional Power System Operations," December 21, 1999.

65 PJM, ISONE Press Release, "ISO New England and PJM Interconnection Propose a Standard Market Design for Wholesale Electricity Markets," March 29, 2001.

66 For a further discussion, see William W. Hogan, "Regional Transmission Organizations: Millennium Order on Designing Market Institutions for Electric Network Systems," Center for Business and Government, Harvard University, May 2000.

67 Ministry of Economic Development of New Zealand, "Inquiry into the Electricity Industry," Report to the Minister of New Zealand, Wellington, New Zealand, June 2000.

without regulators. The latest reform of the reforms was not complete at the end of 2000, but it did give evidence of continuing the process of careful examination of all the pieces and their interdependence. The Government of New Zealand set down principles for reform of the electricity market and development of new regulatory arrangements.[68] These principles could serve as a model for other countries.

The foremost missing ingredient in the New Zealand wholesale market design is a system of long-term transmission rights. A further extension of the New Zealand design would allow for a connection between short-term operations and long-term contracting by providing FTRs. It is straightforward that the monopoly transmission provider must be the first source of transmission rights. These rights might be tradable in a secondary market, but the fundamental definition, initial award, and ongoing provision of the transmission rights must be handled through the transmission provider. Furthermore, the transmission rights must be made compatible with the operation of the coordinated spot market. The special characteristics of the electricity network complicate the definition and provision of long-term transmission rights. The use of FTRs provides a consistent solution that is both theoretically sound and demonstrated in successful applications.

At the end of 2000, there was common agreement that preserving the best features of the existing New Zealand wholesale market design should be a high priority. Furthermore, there was agreement that extending the model to include FTRs would provide an added tool that would provide mechanisms for hedging transmission congestion costs and incentives for long-term investment.[69]

England and Wales

The case of England and Wales presents an exception and a challenge to the argument developed in this paper. The initial reforms in England and Wales in 1990 were highly influential in subsequent developments in electricity restructuring around the world. The signature element of the model was the introduction of the "Pool" by which the system operator managed a coordinated spot market. The principal difference from the British design and the essential ingredients described above was the reliance on a single zone in place of locational pricing to recognize the effects of transmission congestion. The problems created by this exception were managed through a combination of socialization of the congestion costs and a policy of guaranteeing full access to the grid for all generators. In practice, this meant that generators in certain regions would be paid not to generate power when the system was constrained.

The perverse incentives that flowed from this pricing system created a predictable market response that led to a rapid increase in the cost of managing congestion. This could not be sustained, and the policy response was to provide incentives for the National Grid Company to manage the transmission grid, set locational connection charges, and absorb a fraction of the congestion costs. In effect, this approach reverted to the use of a monopoly with price cap

---

[68] Pete Hodgson, Minister of Energy, "Government Policy Statement: Further Development of New Zealand's Electricity Industry," Wellington, New Zealand, December 2000.

[69] Ministry of Economic Development of New Zealand, "Inquiry into the Electricity Industry," Report to the Minister of New Zealand, Wellington, New Zealand, June 2000, p. 61.

regulation in order to provide incentives to counteract the effects of inefficient pricing presented to the market participants. This left problematic incentives for the location of new generating plant, much as in New England, but on balance the system seemed to be working reasonably well.[70] This particular solution would be difficult to transport to another country where multiple interconnected system operators would be found with parallel flows, not just one system operator with a few controllable interconnections. In any event, the locational incentive problems remain and "…the costs of inefficient location can be large compared to the benefits of competition."[71]

The more persistent problem in England and Wales was the concern over the ability of the relatively few large generating companies to manipulate the pool price.[72] Although there were some divestitures of existing generating plants and a substantial volume of new construction, the concern remained that the exercise of market power was a problem. Of course, no market is perfect, and different observers might come to different conclusions about the seriousness of the market power problem in England and Wales. However, this is a value judgment, and the British regulator came to the view that something needed to be done about market power and other features of the market design.

The subsequent argument and analysis took an unusual turn, however, when the conclusion emerged that the very design of the British pool and its coordinated spot market facilitated, even caused, the exercise of market power. The argument arose that somehow the formal application of the economic law of one price made it easier to manipulate the market, and the transparency of the pool reinforced this ability.

Thus arrived the New Electricity Trading Arrangements (NETA) for the market in England and Wales.[73] The new system is complicated, but the essence is simple. Market participants would be required to arrange bilateral transactions at confidential prices. As always, the desire to rely completely on decentralized trading could not be realized. There is still a need for a system operator providing coordination services. Hence, in the NETA design the old day-ahead pool based on a coordinated spot market with a market-clearing price was replaced by a three-and-a-half-hour ahead balancing system with a complex pricing scheme that features a pay-as-bid mechanism with rules intended to penalize imbalances. In effect, the old coordinated spot market with relatively efficient pricing was replaced with a new coordinated spot market with inefficient and obscure pricing.

A complete analysis of the features of this reform of the reform in England and Wales is beyond the scope of the present paper. However, there is substantial support for the view that the NETA reform premise was misplaced:

> "The government believes that the Pool has been biased against coal-fired generators, and that its price-setting rule (all generators are paid the bid of the

---

[70] David M. Newbery, Privatization, Restructuring, and Regulation of Network Utilities, MIT Press, 1999, pp. 210.

[71] David M. Newbery, Privatization, Restructuring, and Regulation of Network Utilities, MIT Press, 1999, pp. 269.

[72] Catherine Wolfram, "Measuring Duopoly Power in the British Electricity Spot Market," American Economic Review, Vol. 89, No. 4, September 1999, pp. 805-826.

[73] For details on NETA see the UK regulator: Office of Gas and Electricity Markets, "Balancing and Settlement Code," March 1, 2001.

> marginal unit) has inflated the level of prices. In practice, many of the perceived problems in the Pool are the result of market power, not the basic design of the Pool, which is capable of sending the right price signals to generators."[74]

Other economic analyses come to even stronger conclusions that the policy does not hold up under logical scrutiny,[75] will increase costs,[76] and should not be followed by the rest of the world.[77]

In any event, this particular reform of reforms is fully supported by the British regulator and was launched in March 2001 after a great deal of preparation and expense. The early days included the expected startup problems. The lights have stayed on, but it is too early to tell much. It will be of great interest to follow the progress of NETA. It is a test of the main argument here. By the analysis above, we would expect the use of inefficient pricing in the spot market to result in greater costs for market participants and substantial unanticipated market behavior. This in turn will produce more, not less, intervention by both the regulator and the monopoly system operator as they then seek to undo what the market has done.

California

The most prominent early death of an electricity reform appears to be a suicide by reckless behavior. At the turn of the millennium, the early promise of the California electricity market reforms unraveled in the cascading collapse of a major market and the worst electricity restructuring policy failure ever seen or even previously imagined. By the end of 2000, a power crisis in California laid bare the dangers of designing a market while ignoring the fundamentals of how power systems operate. A flawed wholesale market and a caricature of a retail electricity market arose in California as the product of a volatile combination of bad economic theory and worse political economy practice.

Bad design outcomes were compounded by bad luck. There had been little addition to generating capacity for more than a decade. Low water reservoirs behind power dams combined with higher natural gas prices and tighter environmental conditions. An unexpected surge in demand from economic growth hit the inefficient market and produced unprecedented price increases. In the event, starting in June 2000 wholesale prices surged and stayed above $150 per MWh while retail prices for the same energy were limited to approximately $65. The system soon fell apart, the lights literally began to go out, and "deregulation" was pronounced dead.[78]

---

[74] Richard Green's "Draining the Pool: The Reform of Electricity Trading in England and Wales", Energy Policy, Vol. 27, No. 9, 1999, p. 515.

[75] Alex Henney, "The Illusory Politics and Imaginary Economics of NETA," Power UK, 85, March 2001, pp. 16-26.

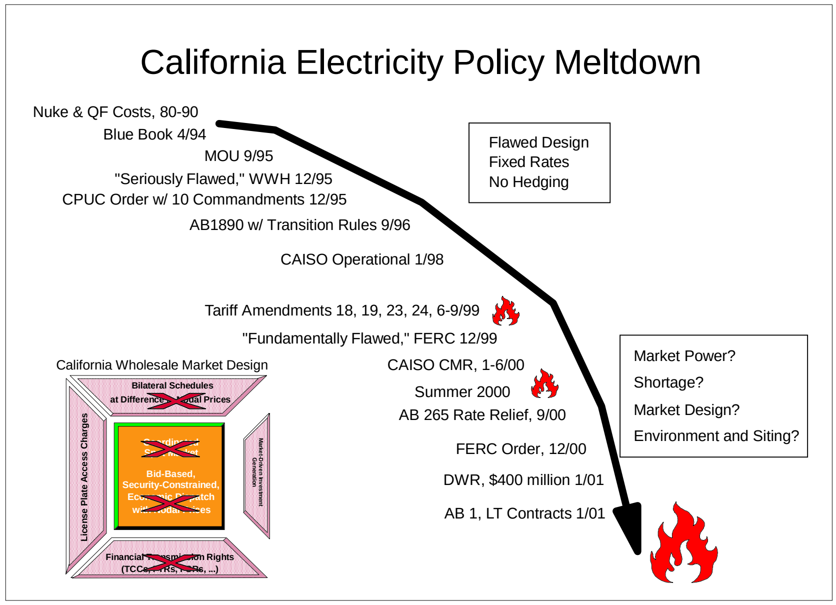
[76] Bower, John and Derek W. Bunn, "Model-Based Comparisons of Pool and Bilateral Markets for Electricity," Energy Journal, Vol. 21, November 3, 2000, pp. 1-29.

[77] Catherine D. Wolfram, "Electricity Markets: Should the Rest of the World Adopt the UK Reforms?", Regulation, The Cato Institute, Vol. 22, No. 4, 1999, pp. pp. 48-53.

[78] For a polished review of the elements of the unhappy combination of events, see Bay Area Economic Forum, "The Bay Area - A Knowledge Economy Needs Power: A Report on California's Energy Crisis and its Impact on the Bay Area Economy," April 2001. (www.bayeconfor.org)

A full investigation of this subject would take us far from the main topic.[79] However, for the subject at hand, California is important because the market was in trouble well before it spun out of control in the summer of 2000. Even without its run of substantial bad luck and exploding prices, the California reform needed reforming, almost immediately.

As the political process took over in 1995, California turned away from the regulator's "Blue Book." Instead, California built its market design on a flawed premise that the inescapable reality of coordination requirements could be ignored or minimized in an effort to honor a boundless faith in the ability of markets to solve all problems. Worse yet, the design of the California market embraced the notion that what little the system operator would do should be done inefficiently in order to leave even more coordination problems for the market to solve.[80] This was an unprecedented experiment with a "seriously flawed" market design that did not work in theory.[81] We now know that it did not work in practice either.



The bad economic theory was a full embrace of the objective of creating a market for middlemen, no matter what the cost. In California, the approach of a coordinated spot market was explicitly rejected in preference to a complicated trading regime as embodied in the Memorandum of Understanding of 1995.[82] The subsequent 'ten commandments' from the CPUC

---

79 For an expanded version of this discussion and further references on the California market design failure, see John D. Chandley, Scott M. Harvey, William W. Hogan, "Electricity Market Reform in California," Comments in FERC Docket EL00-95-000, Center for Business and Government, Harvard University, November 22, 2000.

80 Steven Stoft, "What Should a Power Marketer Want?", The Electricity Journal, 1997, pp. 34-45.

81 William W. Hogan, "A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy," Electricity Journal, December 1995, pp. 26-37.

82 "Professor Hogan can also be read to suggest that the ISO should become the 'pool' by taking schedules which include not just quantity information, but also include price information so that the ISO can select 'the most economically efficient' requests from among the schedules, as if the schedules were bids into the pool. This proposal would essentially re-create the pool in the guise of the ISO. Again, there can be no doubt that the parties intended to foreclose this situation. Indeed, the parties went to great lengths in the MOU to allow buyers and sellers to purchase unbundled transmission rights, to make quantity-only schedules, and not to disclose pricing information to the ISO or subject their transactions to 'economic dispatch.'" Enron et al., "Comments of Enron Capital & Trade Resources, Wickland Power Services, Destec Power Services, inc., Illinova Power Marketing, Inc., Coastal Electric Services, and Electric Clearinghouse, Inc., on the Memorandum of Understanding filed September 11, 1995," dated

attempted to undue these errors,[83] but these commandments were ignored in the resulting enabling law AB1890 and the implementation of the market design with the CAISO and a separate Power Exchange (PX). Given the inevitable requirements for coordination, this produced an expanding collection of arcane rules to prevent what was natural by making the coordination process ever harder to use, all in the interest of supporting separate exchanges and marketers. For example, the CAISO was explicitly precluded from providing a least-cost combination of balancing services.[84] Since the operator still had to provide balancing services, these were required to be inefficient and expensive, in order to create more business for the middlemen. Eventually the CAISO and the PX were operating so many un-coordinated and inconsistent markets for energy and various ancillary services that it was amazing it worked at all.

The compounding failures in the market design accumulated from the market's inception in 1998. Many of the problems that confronted the California ISO and market participants had a common origin in the limitations of the congestion pricing system. California has had essentially a two zone congestion pricing system that was characterized by the existence of considerable intra-zonal congestion, the prospect of additional intra-zonal congestion in the future, and load regions in which high cost generation and transmission investments would be required to meet future load growth. This system has not worked and cannot work in the long-run, because it does not provide generators with the right incentives either with respect to short-run operating decisions or long-run investment decisions.

For example, the constrained-off payment mechanism for managing intra-zonal congestion did not provide generators the right incentives in either the short- or long-run. Generators that were backed down in real time due to intra-zonal congestion received constrained-off payments based on the zonal price. Hence, inefficiently high cost generators would remain in operation, and there was a potential incentive for inefficient entry of new generators requiring additional constrained-off payments. Moreover, there were short-run circumstances in which intra-zonal transmission constraints would create gaming opportunities for individual generators that could schedule transactions in the day-ahead market for which very high constrained-off payments could be extracted in the real-time market.

Amendments 19[85] and 23[86] to the CAISO tariff attempted to reduce the potential for inefficient outcomes under the constrained-off payment mechanism by placing a variety of restrictions on generator choices. Rather than correcting the market design flaws, these amendments addressed market imperfections by adding command and control mechanisms that would likely serve as barriers to efficient generation entry in the case of Amendment 19 and

---

October 2, 1995, and filed with the CPUC, p. 13.

83 California Public Utility Commission, Decision 95-12-063 December 20, 1995, as amended by D.96-01-009, January 10, 1996, Section III, Part 2. See the ten "Principles for Operation of the ISO."

84 William W. Hogan, "WEPEX: What's Wrong With Least Cost?" Public Utilities Fortnightly, January 1, 1998, pp. 46-49.

85 CASIO tariff proposed Amendment 19 Docket No. ER99-3339-000 (New Generator Interconnection Policy), June 23, 1999.

86 CAISO tariff proposed Amendment 23 Docket No. ER00-555-002 (Hourly Ex-Post Price), November 10, 1999.

would lead to an inefficient non-market based intra-zonal redispatch in the case of Amendment 23.

The Amendment 19 policy was not new. Essentially the same type of proposal was in place in New England until October 1998.[87] The policy was defined for similar reasons; namely, to offset the perverse incentives of zonal price aggregation and had the same effect of protecting the incumbent generators. The New England policy was abandoned as inefficient, unfair, and unworkable. As a result, New England revisited the market design issues which created the problem the policy was trying to solve.

At the same time, the CAISO recognized that the pricing system had not drawn forth the necessary level of generation and transmission investment within the transmission constrained areas that include most of the load in California. Hemmed in by its basic design principles, the CAISO sought to remedy the inadequate returns and lack of investment within transmission constrained regions by proposing a process that would govern the distribution of additional extra-market payments to generators within constrained regions (Amendment 24[88]).

Finally, in December 1999, the FERC rejected the ad hoc market adjustments and call for fundamental reform of the zonal congestion management system. "The problem facing the [California] ISO is that the existing congestion management approach is fundamentally flawed and needs to be overhauled or replaced."[89] By the usual standards of dry FERC prose, this was strong language. There then began an intense process to rethink congestion management, and soon the full market design, from first principles.

It was a race against time. Time ran out. When the bad luck arrived in the summer of 2000, California's Comprehensive Market Redesign (CMR) effort was blown back as the explosive combination of variable wholesale prices and fixed retail prices confronted the spark of a suddenly tight market. Bad luck collided with bad policy. The California government intervened with AB265 to impose retail price caps in San Diego, the one region that had moved to a retail market. The FERC issued a series of orders that reflected a view that the problems must largely be solved in California. The state Department of Water Resources jumped in to buy power for the near or soon to be bankrupt utilities who stopped paying their bills. The state then launched a long-term program, beginning with law AB1, to take over or at least play a prominent role in the electricity market. Even those who predicted problems were surprised at the scope and speed of the policy disaster.

The tragic case of California reinforces the basic argument of the present paper. The magic of the market is no sure thing. The details matter. However, the conditions were so extreme in California that even a good market design may not have survived the summer of 2000 and its aftermath. The outcome of all this was unknowable as the summer of 2001 was about to

---

87 See New England Power Pool, 85 FERC Para 61,141 (1998).

88 CAISO tariff proposed Amendment 24 Docket No. ER00-866-000 (Revised Long Term Grid Planning), December 21, 1999.

89 Federal Energy Regulatory Commission, "Order Accepting for Filing in Part and Rejecting in Part Proposed Tariff Amendment and Directing Reevaluation of Approach to Addressing Intrazonal Congestion," Docket ER00-555-000, 90 FERC 61, 000, Washington DC, January 7, 2000, p. 9. See also Federal Energy Regulatory Commission, "Order Denying Requests for Clarifications and Rehearing," 91 FERC 61, 026, Docket ER00-555-001, Washington DC, April 12, 2000, p. 4.

begin. But everything would be different after the experience of such a major failure of market restructuring.

CONCLUSION

The developing experience around the world provides insight into the options and implications of alternative models of access to transmission grids in support of an efficient competitive electricity markets. It is argued from this experience that the central wholesale market design requirement is easy access to a coordinated spot market. There are certain critical functions that must be provided by the system operator. When these functions are organized within the framework of a bid-based, security-constrained economic dispatch with locational pricing, the market has the tools available to deal with the most important network complexities that otherwise confound electricity markets. Furthermore, there must be a close connection between the design of options for market flexibility and the pricing principles for actual use of the transmission grid. If prices closely reflect operating conditions and marginal costs, then market participants can have numerous choices in the way they use the transmission system. However, if pricing does not conform to the operating conditions, then substantial operating restrictions must be imposed to preserve system reliability. Customer flexibility and choice require efficient pricing; inefficient pricing necessarily limits market flexibility. As experience develops, the reforms of reforms reveal just how critical are the details of electricity market design, and how they constrain what can be done.