REGIONAL TRANSMISSION ORGANIZATIONS: MILLENNIUM ORDER ON DESIGNING MARKET INSTITUTIONS FOR ELECTRIC NETWORK SYSTEMS

WILLIAM W. HOGAN

May 2000

Harvard-Japan Project on Energy and the Environment

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

CONTENTS

REGIONAL TRANSMISSION ORGANIZATIONS: MILLENNIUM ORDER ON DESIGNING MARKET INSTITUTIONS FOR ELECTRIC NETWORK SYSTEMS

William W. Hogan[1]
May 2000

"...it is clear that RTOs are needed to resolve impediments to fully competitive markets."[2]

The Commission identified the key role of RTOs in supporting competitive electricity markets. No matter what the name or governance structure, there must be an organization responsible for key coordination activities needed to make a market work. This is not an option, and the market cannot solve the problem of market design. Hence the most important part of the Commission's Millennium Order is the description of a workable framework for competitive electricity markets. A competitive electricity market can be the 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. The central 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 tools are available to deal with the most important network complexities that otherwise confound electricity markets. Once done, many of the other problems in the electric network would either disappear or would be greatly simplified.

INTRODUCTION

The Federal Energy Regulatory Commission has addressed a wide range of issues in its analysis of and orders for the design of Regional Transmission Organizations (RTO).[3] To signal its importance, the Commission assigned the RTO Order the millennium number. This Millennium

---

[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 Law and Economics Consulting Group in Navigant Consulting, Inc. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. This revises the author's earlier paper of August 1999, submitted as part of the RTO Notice of Proposed Rulemaking. The author is or has been a consultant on electric market reform and transmission issues for American National Power, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, Calpine Corporation, Commonwealth Edison Company, Detroit Edison Company, Duquesne Light Company, Electricity Corporation of New Zealand, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., 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, TransÉnergie, Transpower of New Zealand, Westbrook Power, Williams Energy Group, and Wisconsin Electric Power Company. 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://ksgwww.harvard.edu/people/whogan).

[2] RTO Order, p. 115.

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

RTO Order covers a great deal, and says a great deal, going a long way in defining the minimum characteristics and functions that must be provided in support of an open, reliable, and robust competitive electricity market. As is perhaps inevitable in the process with a document intended to address so many contentious issues, however, some of the most important ideas of the RTO Order call out for further development and an unmistakable commitment. The key ideas are there, but they need greater emphasis in the design and greater priority in the schedule so as not to be foreclosed by the unanticipated consequences of seemingly unrelated decisions.

While many other issues such as the form of ownership, governance structure, regional boundaries, independence of market participants, or incentive regulation can be significant, they should be fashioned nevertheless only in the service of well-designed market institutions. The fundamental guiding principle of RTO design should be to serve the public interest. A competitive electricity market can be the vehicle for pursuing that 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. 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 to be discovered through the magic of the market. After all, the whole point of moving to greater reliance on markets is the belief that the market participants will respond to incentives, fast. But markets with poorly designed institutions will give the wrong incentives. The mistakes, once made, will not be easy to fix.

As discussed below, the Commission's RTO Order contains the elements of an efficient market design built around a system operator that coordinates a spot market. This efficient, coordinated spot market is the only design we know of that is both internally consistent and actually works. More than anything else, implementation of this efficient design, and all that it entails, should be first on the agenda. Extensive delay in adopting the design will prove costly. Without an efficient spot market and its ease of access, the problems of discrimination will persist. Without an efficient spot market and its consistent incentives, operational problems will force system operators to impose administrative command-and-control procedures that defeat the purposes of the market participants. Without an efficient spot market and the associated transparent spot prices, it will be much more expensive and difficult to arrange balancing and settlement for the increasing number of retail access programs in the states. Without an efficient spot market and the associated locational prices, there will be no way to define a workable system of transmission rights, no way to stimulate investment in transmission by market participants and, therefore, no way to avoid complete reliance as of old on monopoly decision-making and investment. Without an efficient spot market designed through the RTO process, the Commission will inevitably face exactly the same design questions in the development of the unavoidable transmission loading relief procedures, where important market structure decisions are being made under the name of reliability.[4]

THE ESSENTIAL MARKET INGREDIENTS

There are a few critical ingredients in the design of an efficient competitive electricity market. Given the highly interconnected network, it is clear that some aggregation to regional transmission organizations would be necessary. The issues involved in the development of RTOs are important, but the discussion has tended to focus on ownership and governance questions.

---

4 RTO Order, pp. 69-70.

These can distract from the more difficult but more fundamental consideration of the rules for market operations within and across regions in network systems.

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. 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. At this stage of the national debate, it is no longer necessary to repeat the analysis to support the conclusion that the complex network interactions in an electric grid require that there be an entity that can provide certain critical coordinating services.[5] 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 the same laws in order to manage the flow of power within the limits of the transmission system.[6] 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 they 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 the RTO Order and its implementation should be to require good design for the functions of the system operator.

The example of energy balancing illustrates the point. At all times, the system operator must coordinate increases and decreases in dispatch to maintain aggregate real power balance. And

---

[5] 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.

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

when the transmission system is constrained, the system operator must arrange the redispatch to ensure that the free flow of power stays within the security constraints of the system. Hence, energy balancing and congestion management are inextricably intertwined. In performing these functions, the choices faced by the system operator could be summarized under three questions.

- **Should the system operator offer an economic (re)dispatch service?**

  Market participants can schedule their proposed use of the transmission system, or make offers to buy and sell energy. The energy balancing redispatch service amounts to buying and selling power at the margin. In a competitive market, producers could provide bids to generate more or less. The selection of the least cost combination is the natural criterion. This is familiar in the industry as economic dispatch. The alternative would be to define a set of administrative procedures and rules for system balancing that would purposely ignore the information about the costs of running particular plants. It seems more natural that the operator consider customer bids and provide an economic (re)dispatch.

- **Should the system operator apply marginal cost prices for power provided through the dispatch?**

  Economic dispatch is consistent with the competitive market outcome, along with the natural market-clearing prices. Under an economic dispatch for the flexible plants and loads, it is a straightforward matter to determine the locational marginal costs of additional power. These marginal costs are also the prices that would apply in the case of a perfect competitive market at equilibrium. In addition, these locational marginal cost prices provide the consistent foundation for the design of a comparable transmission tariff. Any other pricing system would be inconsistent with the self-enforcing property of market-clearing prices. Any other pricing system would create perverse incentives which would drive the system operator away from the market and towards increasingly restrictive command-and-control type rules.

- **Should generators and customers be allowed to participate in the economic dispatch offered by the system operator?**

  The natural extension of open access and the principles of choice would suggest that participation in the coordinated balancing market offered by the system operator should be voluntary. Market participants can evaluate their own economic situation and make their own choice about participating in the operator's economic dispatch or finding similar services elsewhere. Any other rule would require some form of discrimination, either to prevent participation by some in the balancing market, or to compel participation by others.

The simple answer to the three question is to "Just Say Yes." Then, as discussed at greater length below, the best approach would be 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 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.[7]

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. Long-term transmission rights could be defined as financial rights to the difference in locational prices, thereby avoiding the impossible problem of defining a set of so-called "physical" transmission rights that would be adequate for managing the use of the grid.

These benefits, created through a coordinated spot-market using a bid-based, security-constrained economic dispatch with locational or nodal prices, are developed further in an appendix on the economics of a competitive electricity market. The theory of the case is by now well supported by practical experience. The main ingredients exist in many parts of the world, and the combined package has been operating successfully in the Pennsylvania-New Jersey-Maryland Interconnection (PJM) for more than two years.[8] The same model has been adopted in New York,[9] and embraced as a reform in New England.[10] Likewise, the problems that arise when we do anything else are apparent in various experiments where putative simplifications produced predictable problems.[11]

The final name we give to the system operator is not as important as recognition of its essential functions and how they should be performed. Anticipating the discussion of alternative organizational arrangements, recognize that this may be a hands-on system operator, one who actually physically controls the power plants. Or the degree of control may be more indirect. The system operator could control the computers that control the dispatch. Or the system operator could give instructions to the control area operators who control the computers that give instructions to the plants. Or the system operator could set the rules and run the computers that give instructions to the operators who give instructions to the computers that give instructions to the plants. Someone must provide the essential coordination function, and it is easiest to think of this role as being

---

7 Larry Ruff, "Competitive Electricity Markets: Why They Are Working And How To Improve Them," National Economic Research Associates, May 12, 1999.

8 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.

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

10 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.

11 William W. Hogan, "Restructuring the Electricity Market: Institutions for Network Systems," Center for Business and Government, Harvard University, April 1999, pp. 38-42, available from the author's web page.

coincident with the role of the system operator. But the organizational context within an RTO framework is secondary, even a distraction. What is important is that the coordination function be provided, and provided to meet the requirements dictated by the electrical network while supporting the operation of a competitive market.[12]

A close reading of the Millennium Order can be used to support the argument that the Commission knows all this.[13] The Order sets forward a broad framework for electricity restructuring in support of competitive markets through its requirements for minimum "characteristics" and "functions" of RTOs:

> "...the four minimum characteristics for an RTO: ...
>
> (1) independence from market participants;
>
> (2) appropriate scope and regional configuration;
>
> (3) possession of operational authority for all transmission facilities under the RTO's control; and
>
> (4) exclusive authority to maintain short-term reliability."[14]

In addition, there are eight minimum functions that an RTO must perform.

> "...an RTO must:
>
> (1) administer its own tariff and employ a transmission pricing system that will promote efficient use and expansion of transmission and generation facilities;
>
> (2) create market mechanisms to manage transmission congestion;
>
> (3) develop and implement procedures to address parallel path flow issues;
>
> (4) serve as a supplier of last resort for all ancillary services required in Order No. 888 and subsequent orders;
>
> (5) operate a single OASIS site for all transmission facilities under its control with responsibility for independently calculating TTC and ATC;
>
> (6) monitor markets to identify design flaws and market power;
>
> (7) plan and coordinate necessary transmission additions and upgrades; [and]
>
> (8) ... ensure the integration of reliability practices within an interconnection and market interface practices among regions."[15]

From the perspective of design of institutions, the most important theme running through the Commission's discussion of these characteristics and functions is the prominence of markets as the means for achieving the many goals of the RTO. The Commission recognizes the importance of organizing and using a market, and builds on this core idea. But it still requires a close reading to

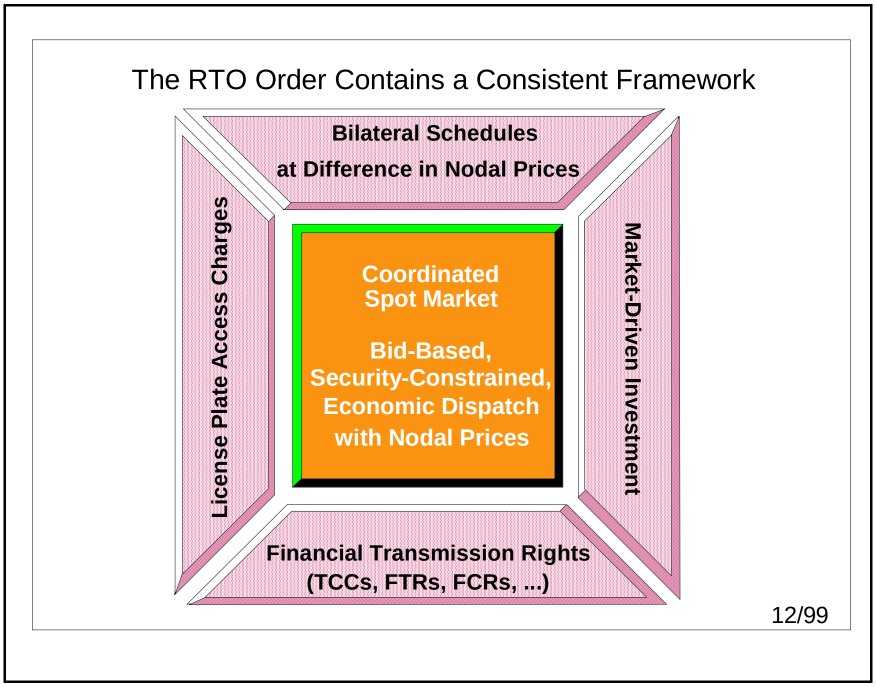
---

12 RTO Order, pp. 268-269.

13 RTO Order, pp. 632-643.

14 RTO Order, p. 152.

15 RTO Order, pp. 323, 495.



recognize the details and organize what is in the public interest in supporting a well-functioning competitive market.

The close reading finds the basic components of the essential competitive market structure in the proposed requirements for RTO. The key element is in the recognition of the importance of a coordinated spot market. In the RTO Order this appears principally in the discussion of the balancing market. The Commission has answered "Yes" to the three questions posed above. In particular, the Commission recognizes that "[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."[16] 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 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."[17]

Further, the Commission 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."[18] 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."[19]

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

---

[16] RTO Order, p. 635.

[17] RTO Order, pp. 332-333. See also p. 382.

[18] RTO Order, p. 332. See also p. 743.

[19] RTO Order, p. 423. See also p. 715.

giving rise to the creation of financial transmission rights.[20] 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.[21]

These are the most important elements. These define the functions of the essential system operator. There 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. Putting these rules in place should be of the highest priority, and everything else in the design of market institutions should be examined as to how it supports or contradicts this basic structure.

INSTITUTIONS AND 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. There must be a system operator. The only open questions are with the rules the system operator will apply and the governance of its activities.[22] The development of Independent System Operators (ISO) 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 RTO Order summarizes the range of debate in the United States about the Commission's authority to mandate membership in an ISO, the need for such mandates, the search for alternative models, and the possibility that ISOs might be only a transitional arrangement. For example, there have been suggestions that ownership of the wires (a Gridco) combined with system operations (an SO) could produce an independent transmission company (a Transco) that would be different from an ISO, or an alternative that might be precluded by an ISO. Both the Gridco and the Transco might be described as an independent transmission company (ITC).[23] For the present discussion, however, it is important to maintain the distinction of whether or not the system operator is combined with the ownership of the wires. Any or all of these models could become the eventual form of the RTO.

---

[20] RTO Order, pp. 382-383.

[21] RTO Order, p. 524.

[22] William W. Hogan, "Independent System Operator: Pricing and Flexibility in a Competitive Electricity Market," Center for Business and Government, Harvard University, February 1998.

[23] For example, a version of the Transco model, as an independent transmission company proposed by Commonwealth Edison et al. received initial approval in Federal Energy Regulatory Commission, "Order Granting in Part Petition for Declaratory Order, Providing Guidance and Accepting Amendment for Filing," 90 FERC 61, 192, Washington DC, February 24, 2000. The filing details the market design and principles of operation of the ITC along the lines outlined above.

There is a continuing debate about the best model for organizing coordination and control of the transmission system.

Transco. An independent company that combines ownership of the grid and responsibility for system operations in managing the use of the grid. May be a for-profit or not-for-profit entity. (National Grid Company in England and Wales.)

Gridco. An independent company that owns the grid but does not have responsibility for operating the system. Works in conjunction with a system operator. May be a for-profit or not-for-profit entity. (GPU PowerNet in Victoria, Australia)

ISO/PX. An independent system operator with restrictions to allow for separate operation of a power exchange. (California ISO and PX.)

ISO. An independent system operator that has responsibility for managing use of the grid and coordinating the spot market. (Pennsylvania-New Jersey-Maryland Interconnection, PJM.)

TLR. The institution for coordinating transmission loading relief across regional system operators. (NERC Security Coordinators in the U.S. Eastern Interconnection.)

A symptom of the confusion over the rules for a competitive market is in the parallel activities devoted to the discussion of ISOs, the Commission's OASIS system for transmission scheduling,[24] the North American Electric Reliability Council (NERC) security coordinators for transmission line loading relief, and the Commission's earlier Capacity Reservation Tariff (CRT) proposals.[25] Although these packages tend to be discussed in isolation, there is substantial overlap in that they all provide alternative approaches for the same core problem: rationing use of scarce transmission capacity. Furthermore, the approaches tend to be mutually inconsistent: some ISO models include bid-based economic dispatch; OASIS (in practice, if not in theory) is built around the flawed contract-path model; NERC's tagging rules and line loading relief procedures struggle to undo the contract path fiction, in order to deal with power flow realities and the commercial complications of administrative curtailments; the CRT would move all the way to a point-to-point reservation system with economic rationing.[26] We can't do all of these at the same time. And the

---

24 Open Access Same time Information System (OASIS), FERC Order 889, Final Rule, Washington, DC, April 24, 1996.

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

26 Michael Cadwalader, Scott Harvey, William Hogan, and Susan Pope, "Coordination of Congestion Relief

attendant problems of coordinating trade across regions may be some of the most vexing for the regulators and the competitive market.

The unwelcome news for the regulators is that the hard problem of allocating scarce transmission capacity is made much more difficult by the move to competitive markets. In effect, we have taken the black box of the vertically integrated industry, opened it, and unbundled control of the various gears. In order for the system to work, however, the gears not only have to turn -- they have to mesh. This is especially true in the very short-run, as we move closer to real-time operations.

Everyone wants non-discrimination and the maximum possible degree of flexibility for market participants. But to provide this flexibility, and make sure the gears mesh, it will be necessary to align the incentives of the participants with the success of the overall market. Either the incentives must match the system realities, or the pricing and access rules of necessity will be more restrictive and dictate customer choices. Furthermore, the role of the system operator inevitably will encompass both reliability and commercial issues. The supposed distinction between reliability and economics is a mirage that will provide no comfort in practice. The nature of the electric grid dictates that decisions motivated by reliability concerns will have substantial commercial impacts, especially when the system is constrained and the decisions matter most. The only issue is the degree to which we will be explicit about the interaction between reliability and economics, in order to improve both efficiency and transparency.

The Commission refers to the range of models as regional transmission organizations, a term intended to encompass many models. Here we focus on the implications for competition in generation, and the rules for the wholesale market. With a focus on market institutions needed to support competitive markets, a critical summary of the debate over transmission models highlights the importance of system operations and real-time dispatch.

Transco

The Transco model as defined here emphasizes the combined responsibility for ownership of the wires and conduct of system operations. A Transco is a single regional entity that owns and operates the transmission system, but is independent of generation and load. The emphasis on control of system operations isolates one of the key elements that define the relationship with the design of institutions for the competitive market.

It seems only natural that ownership of an asset should imply control of its use. However, unlike most other markets, this link between ownership and control of operations is literally not possible for an interconnected electric network. Absent a single entity for the entire grid, there is no avoiding the necessity of having operations controlled at least in part by someone else. Hence, for electricity, setting the rules for how you use your own asset is unavoidable. The complications do not disappear through a simple change of ownership and governance. The key function of system operations, as illustrated above in terms of energy balancing, must be performed within a framework that supports the public interest in a competitive market. Embedding the responsibility for execution of this function within a Transco would not change the necessity to define the rules according to objectives that transcend those of the Transco alone.

Across Multiple Regions," Center for Business and Government, Harvard University, October 7, 1999.

The leading early proposals call for regulated for-profit entities.[27] In part, the motivation for creating a Transco is to exploit the incentive effects of the profit motive. Presumably the profit opportunities would provide inducement for improved operations and market responsive investment. At a minimum, with ownership of significant assets, there is an argument that regulators would have greater leverage in controlling the performance of Transcos.

The strongest claims are that the profit motive is all that would be needed, and with appropriate incentive regulation the Transco could be left to devise its own rules for transmission access, operations and detailed pricing. By this argument, mere establishment of a for-profit Transco would dispense with the difficulties of evaluating the pricing and access rules for transmission and system operations.[28] Apparently through some type of incentive regulation, an independent Transco would support a non-discriminatory, competitive electricity market that meets the Commission's public policy goals. While this may be a theoretical possibility, there is no known system of incentive regulation that could achieve this result. The difficulties to be overcome would begin with the same set of problems that complicate the process of setting the rules for system operation. At the core is the uncomfortable reality that there is no simple definition of the output of the transmission system. Efficient transmission is far more than electric throughput -- it is a complex service with many dimensions and substantial network interactions. Were this not true, there would probably be no need for a system operator in the first place.

The Federal Trade Commission (FTC) has identified a flaw in the argument that a Transco would necessarily have the right incentives to support a competitive electricity market.[29] A critical problem appears in the possible substitution between transmission and generation. We learned from many years of utility investment planning analysis that there is always a tradeoff between generation and transmission solutions when the system becomes constrained. It follows then that ownership of the wires and control of system operations (which means controlling the dispatch) would create an inherent conflict of interest for a Transco, with incentives to tilt operations to induce or preclude investment so as to benefit the Transco.

Furthermore, the assumption that it would be an easy matter to set the proper incentives for a Transco, incentives sufficient to leave to management the choice of rules and procedures for system operations, runs counter to the whole notion of electricity restructuring and greater reliance on the market. If we were so confident that we knew how to regulate such monopolies, then there would no need for restructuring and unbundling.[30] Quite to the contrary, it is a difficult matter to set such incentives. There is always a tradeoff between alternative compromises, and the design of a Transco model must confront its limitations.

---

[27] For example, see the initial public announcements by Entergy, Northern States Power, First Energy, and Commonwealth Edison.

[28] For example, see Curt L. Hébert, Jr., "Moving the RTO Debate," The Electricity Journal, March 1999, pp. 20-23.

[29] Federal Trade Commission, Before Public Service Commission of the State of Mississippi, Docket No. 96-UA-389, August 28, 1998. Federal Trade Commission, Before the Federal Energy Regulatory Commission, In the Matter of Entergy Services, Inc., "Comments of the Staff of the Bureau of Economics of the Federal Trade Commission," Docket No. EL99-57-000, May 27, 1999.

[30] Lawrence J. Spiwak, "You Say ISO, I Say Transco, Let's Call the Whole Thing Off," Public Utilities Fortnightly, March 15, 1999.

An independent Transco may be attractive, and it could be the next stage or the end stage. But the very complexities that dictate the need for a system operator mean that it will not be an easy matter to structure the rules for system operations, nor would it be easy to structure incentives for a monopoly to discover the rules on its own. Providing appropriate incentives for the transmission system is a major difficulty in restructured electricity systems around the world. Some problems might be different under the different transmission models, such as the approach to providing incentives for grid maintenance and expansion,[31] but all the puzzles about the operating rules would appear again in this new guise. Hence, it is not likely that the Transco incentives could be developed so easily as to leave design of the system operation rules and pricing to the Transco monopoly alone. Somewhere in the company would be a system operator that must be "ring fenced" from the rest of the corporation, to have its own independent rules and pricing structures that support the public interest in a competitive market, not only the private interests of the monopoly Transco.[32] The Commission will face the task of evaluating and approving the rules for pricing and access.[33] And this applies to the not-so-independent Transcos that are embedded in the vertically integrated utilities, as well as to new independent Transcos that might be divested from the utilities.

Other problems arise when we consider further the Commission's requirements for independence of market participants. Presumably the Commission would support market-driven investments in increased transmission capacity. The form of these investments would have to be defined, but the supposed advantages of combined ownership and control in a Transco might be compromised. Or the independence of the Transco from market participants would be foreclosed. In Australia, for instance, there is a strong interest in promoting merchant-based transmission investments.[34] No sooner had Australia created a small niche for a market-based transmission investment, than construction began on just such an expansion. This is the 180 MW Direct Link project connecting the regional electricity markets in Queensland and New South Wales.[35] The ownership of the line is separate from the control of operations, which will be managed through the Australian ISO. A similar project has been proposed to connect the New England and New York markets via a cable under Long Island Sound.[36]

---

[31] William W. Hogan, "Transmission Investment and Competitive Electricity Markets," Center for Business and Government, Harvard University, April 1998.

[32] Fiona Woolf, Cameron McKenna, comments on Panel 3, "Regulation, Governance, and Independence," FERC Public Conference Concerning the Commission's Policy on Independent System Operators, April 16, 1998. Examples with careful attention to the market rules by which the operators within the Transco pursue dispatch and pricing can be found in the National Grid Company of England and Wales or Transpower in New Zealand.

[33] William L. Massey, "Policy on Regional Transmission Organizations: Five Pitfalls FERC Must Avoid," The Electricity Journal, March 1999, pp. 13-19.

[34] National Electricity Code Administrator, "Entrepreneurial Interconnectors: Safe Harbour Provisions," Transmission and Distribution Pricing Review, Australia, November 1998.

[35] The Direct Link project is a merchant transmission line in Australia developed by TransÉnergie. Further details on the Direct Link project can be found on the TransÉnergie web page at www.transenergie.com.au. The author has been advising TransÉnergie in considering merchant transmission investments in the United States.

[36] TransÉnergie, "Petition of TransÉnergie U.S. Ltd. for Order Accepting Tariff for Transmission Interconnector and Granting Related Authorizations and Waivers," FERC Submission, Washington DC, October 1, 1999.

In the end, therefore, it is unlikely that the Transco would avoid any of the conceptual and design challenges that must be addressed in creating an ISO. In this sense, it would be a mistake to cast a Transco as an alternative to an ISO in any way other than the formalities of organizational charts and governance. The same questions that appear in the specification of the rules for the market, for access and pricing, would appear in establishing the rules for the Transco. The main problems cannot be avoided. In effect, the result is likely to be a de facto ISO within the de jure corporate structure of a Transco. Or we could think of a Transco as an ISO that acquires ownership of the wires.

If we are not relying on the profit incentive alone to produce the rules for system operations, then other approaches to the Transco model might capture some of the benefits of better coordination of transmission investment and wires maintenance combined with an understanding of the needs of system operations. Here the large public power authorities in the United States provide an alternative model with non-profit organizations.[37] This is an old debate, with strong views and conflicting evidence. In the choice between the for-profit and the not-for-profit model, it may be the other details on regional coverage, legal restrictions on the transition, and the model for market operations would be more decisive.

Finally, a major hurdle for the widespread embrace of the Transco model in a large country like the United States would be in creating transmission companies that match the regional requirements of system operations. This is easy in New Zealand where there is already a single transmission owner. It would be much more difficult in the United States, Japan, or Europe with their large interconnected systems. The combination of system operations and ownership of only some of the wires might be much more problematic from the perspective of the owners of the other interconnected wires. Either these other transmission owners must surrender control of operations, foregoing all the presumed benefits of the Transco model, or operations must be balkanized to follow the pattern of wires ownership. If it is not an easy matter to change the patterns of ownership of the wires, therefore, reliance on the true Transco model, which combines complete ownership and operational control in one entity, would substantially hinder the development of integrated markets with broad regional coverage. In the end, the balkanized regional Transcos would not really be able to decide how to operate their systems, because the failure to capture all the regional interactions would leave to the transmission loading relief procedures the de facto definition of operating rules in the face of system constraints.

Gridco

A Gridco is a regional entity that owns transmission wires and is independent of generation and load. As defined here, the Gridco is not responsible for controlling use of the system, and must be paired with a system operator. Many of the advantages of the Transco model would apply to the Gridco approach, but without all of the problems.

Control of operations by an ISO is compatible with the Gridco model. The rules for access and pricing would be the same as under the regime where traditional utilities own the grid that is under the control of a separate system operator. The distinction of the Gridco is that

---

[37] Frank McCarmant, Vincent Tobin, and Stephen Pelcher, "Uncrossing the Wires: Transmission in a Restructured Market," The Electricity Journal, March 1999, pp. 24-35.

maintenance and expansion of the grid could be the responsibility of the Gridco, which is also independent of generation and load.

As with Transcos, the leading proposals call for regulated for-profit entities, such as the strategies embodied in the public announcements of New England Electric System in New England or General Public Utilities in Pennsylvania-New Jersey. In both these cases, the wires company is separated from ownership of generation and from system operations. However, in these cases the companies still own distribution wires, and are not strictly transmission companies as understood in the pure Gridco model. Sharper examples of a pure Gridco approach are found in the transmission companies in Australia such as GPU PowerNet in Victoria, which owns and maintains the transmission wires, but does not own distribution systems and leaves system operations to the independent system operator in the National Electricity Market Management Company (NEMMCO).

The arguments of the large public power authorities apply as well for a mix that includes non-profit Gridcos. The incentives for the Gridco, which would own significant assets, would be similar to those of the Transco, but without the conflicts of interest in operations identified by the FTC.

Because of the separation from operations, regional coverage of the Gridco ownership of the wires need not and probably would not coincide with the regional coverage of system operations. This would be a great simplification compared to the Transco model. It would allow an evolution of Gridcos, with different models, without confronting the complications of balkanized operations. The developing Gridcos would be able to manage transmission investment and maintenance, with appropriate incentive regulation. Providing the proper incentives would not be easy, but it would be possible to pursue market incentives with fewer difficulties in isolating the control of system operations. Investments need not be limited to those of the Gridco. There could be a complex mix of transmission investments by existing utilities and other market participants, either through the Gridcos or in competition with the Gridcos.

Power Exchanges

The debate launched in the California restructuring spawned separate institutions for the operation of the spot market through a power exchange (PX) and control of system operations through an ISO. Here the independent system operator functions in conjunction with a separate and distinct power exchange responsible for market operations, with separate rules and pricing for each. In this case, neither the ISO nor PX owns transmission lines.

The viability of the distinction between the functions of market operations and system operations depend on the time horizon and the relative importance of network interactions. For the short-run, the two functions are difficult (impossible) to separate.[38] Over the short-run, maintaining a distinction between the ISO and the PX requires creation of complex rules to restrict the system operator. It is well recognized that if the system operator performs its functions through use of a voluntary, bid-based, security-constrained, economic dispatch, following the principles power systems have used for decades, the separate power exchange would have little to do other than

---

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

arrange accounting settlements. Hence, in California, where we find a model with this formal attempt to separate the spot market from system operations, the design precludes the ISO from pursuing an economic dispatch and segments interdependent functions, reducing options and increasing costs. The results with this approach have been problematic, at best, and a number of initiatives are underway that will, in effect, undo the artificial separation of markets.[39]

Restrictions on ISOs reappear in various proposals elsewhere that limit the use of economic dispatch and transmission coordination, assuming that the complex interactions can somehow be internalized in a market, even without a formal power exchange.[40] Inevitably these approaches reduce transmission capacity, socialize costs and add to the complexity of real operations.

Note that the issue here is not the emergence of power exchanges to serve the needs of the market. There may be a demand for such services beyond what the system operator can provide. There should be no restrictions on the creation of power exchanges. The issue, rather, is whether there should be restrictions imposed on the ISO in order to create more business for independent power exchanges.

When trading power at a particular location, power exchanges would not confront the all the problems of network interactions. And over horizons where network interaction might not be as important, the advantages for integration of the power exchange and system operations would be reduced. For example, in Norway there is a market which functions as a power exchange separate from the ISO, for trading of contracts and establishment of schedules. But in the end, the true real-time spot market in Norway is the final regulation market administered by the ISO. As the Commission has observed, when the real network interactions come to the fore, there must be a balancing market. In effect this is a coordinated spot market, and it should be integrated in the functions of the system operator rather than artificially separated at the cost of substantial increased complexity and further compromise of the basic market principles.

Independent System Operator

The independent system operator provides a dispatch function that coordinates the spot market. As discussed below, the basic ISO model has been extensively developed in its various incarnations around the world.[41] The ISO does not own transmission lines. If there is a separate entity called a Power Exchange, it does not have responsibility for coordinating the spot market and transmission usage. The PX may handle bidding and settlements, such as with Electricity Market Company (EMCO) in New Zealand. But in New Zealand the real-time dispatch implementation falls to Transpower, which is the de facto ISO. In many cases, there is no separate PX with any special status, as for example in the PJM ISO, the New England ISO, the Australian NEMMCO,

---

39 For example, see Frank Wolak, Robert Nordhaus, and Carl Shapiro, "Report on the Redesign of Markets for Ancillary Services and Real-Time Energy," Market Surveillance Committee of the California Independent System Operator, March 25, 1999.

40 For example, this is the proposal for the Midwest Independent System Operator.

41 William W. Hogan, "Independent System Operator: Pricing and Flexibility in a Competitive Electricity Market," Center for Business and Government, Harvard-Japan Project on Energy and the Environment, Harvard University, March 1998.

and so on.

The services provided by the ISO are complex and interconnected. It is a challenge to find the best mix of unbundled activities and associated pricing rules. The key is to match the degree of customer choice with the pricing incentives. Where customers have flexibility, such as between spot market transactions and bilateral transmission scheduling, it is important to get the prices right. There are many models, each with its own nuances. But there is a core model built on the basic economics of electricity systems, as discussed above.

The appropriate size and regional coverage of the ISO depends on many factors, including the degree of coordination required across the entities in arranging for transmission loading relief. The ISO model is fully compatible with the creation of independent Gridcos, and enjoys the advantage over a Transco that the regional coverage of the single ISO does not have to match that of possibly multiple Gridcos.

Transmission Loading Relief

In large interconnected grids with multiple areas under separate control, regional system operators must coordinate use of the transmission grid. Transmission loading relief (TLR) is required when system constraints would be violated. The rules for inter-regional coordination interact strongly with the pricing and access rules within the regions. The experience in the United States has been that there has been too little in the way of coordination of the TLR rules with the requirements and expectations of the developing markets.

In the United States, the North American Electric Reliability Council (NERC) filled the vacuum in developing a mechanism and institutional framework for TLR. However, the institutional design limits imposed or assumed by NERC required non-market mechanisms for curtailing transactions. In effect, the NERC approach embraced the fallacy of the separation of markets and reliability, assuming that it would be possible to have reliability rules that would be either unimportant or neutral in their commercial effects. The resulting TLR system was cumbersome, reduced real transmission capacity, and had severe impacts on the market, contributing to problems in the Mid-West during the summers of 1998 and 1999 that produced $7000/MWh transactions.[42]

With TLR integrated in the market, prices and bids would matter. The Commission directed NERC to develop more economic systems. There are alternative market mechanisms available in principle. For example, the PJM system proposed implementing the first consistent market mechanism for managing TLR by allowing participants to choose to pay for congestion. This is a separate topic still under active consideration in the United States. The lesson is that the TLR rules must be developed to be consistent with the institutions of the electricity market.[43] The same problems that appear in TLR can be found in the coordination across the seams between regional ISOs. The reliability driven concerns for use of the transmission system cannot be separated from system operations or from the activities of the market, at least in the short-run.

---

[42] 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.

[43] RTO Order, p. 380.

Network interactions are strong, and the same forces that create the need for an ISO drive the need for market driven methods of adjusting use of the transmission system.

MITIGATING MARKET POWER

The essential ingredients of the basic market design presume that there is competition in generation. That assumption stands at the core of electricity restructuring and is the fundamental requirement for moving to a market. The term "competition" means different things to different participants in the market. The list includes a (very) large number of generation suppliers and customers who have the ability to choose among suppliers at will. Furthermore, there would be a (very) large number of customers and suppliers would be able to refuse customers at will. Finally, regulators would want full non-discrimination and comparability.

Likewise, the textbook case for a competitive model supposes a set of limiting assumptions that follow from the assumed structure. Electricity would be a commodity business with no entry costs or barriers. The suppliers would be price-takers with little or no ability to influence the market equilibrium. The same would apply to the many price-taking customers. There would always be a market equilibrium that would fully exhaust arbitrage opportunities under according to the law of one price.

Unfortunately, the limiting conditions of competition between and among generators and customers do not hold in practice. The competitive model is idealized, and there really are no pure cases in the electric market or anywhere else. In reality there are many approximations, and the hope is not that there will be a perfectly competitive market but that there will be at least workable competition. The definition of workable competition is not precise, but the issues are important. The gap between the perfect and the real has many dimensions. In the real world economic dispatch is always imperfect and the system constraints are soft. Investments can be lumpy, especially in transmission and this creates market power and natural monopoly. Almost by definition, market participants seek price advantages whenever they can and both energy and ancillary services can have dominant suppliers.

All these gaps between the real and the ideal can have implications for operation of the market and pricing. It is when the deviations from workable competition move too far towards the exercise of market power that we face the most serious policy problems. And in the real world of electricity, it is not hard to find instances of market power. A full treatment of the issues raised by market power is beyond the scope of the present discussion. However, it is useful to summarize an argument about the implications of market power for the design of market institutions.

Because of its importance, regulators have been concerned with detecting and mitigating market power. The policy response has been a mix of:

- Market monitoring and reporting. (UK, CA, PJM, NY, NE)
- Market power mitigation through divestiture and contracting. (UK, CA, IL, NY, PJM)
- Market power mitigation through market design. (UK, CA, Australia)

Market power monitoring is a among the responsibilities that have been assigned to ISOs, and will probably be a major activity of RTOs. However, the focus here is on market design and the adaptations of market institutions to mitigate the impacts of market power. We consider the case of energy generation, and the debate over the use of efficient pricing mechanisms. There are competing approaches to energy pricing and mitigation of market power in generation.

One approach is to follow efficient pricing principles and leave market power mitigation to other targeted methods (e.g., PJM, NY). Here the theory is that efficient competitive design is the target and appropriate for the long run, with the transitional problems of market power best met by carefully targeted mitigation efforts rather than comprehensive modifications in the market design.

The alternative view is to adopt fundamental changes in the market design and pricing approaches to limit the impact of market power and supplement other targeted methods (e.g., UK, CA, Australia). This raises the problem of making a transition away from the market modifications towards an eventual competitive market model.

Some market design modifications and associated pricing methods might be effective in mitigating market power. For example, one such possibility would be to rely on monopsony purchasing to counterbalance the power of a monopoly seller. However, this well-known theoretical approach would violate the principle of non-discrimination by requiring some rationing among buyers and would be difficult to implement without creating more problems. The use of countervailing market power, therefore, does not seem promising.

The more common approach to market design modification has been to reject efficient pricing based on the competitive model and adopt a blend of socialization and averaging of costs. In particular, it is argued that efficient locational pricing by definition creates small isolated markets where it would be easier to exercise market power. Hence, as the argument goes, in the face of significant market power in generation, zonal aggregation and socialization of costs should replace efficient locational pricing. This is the conventional wisdom found in many policy debates. This conventional wisdom is often unexamined. It is often wrong.

The importance of the conventional wisdom in some places is that it has had deep and pervasive effects on market design and operations. In California, for example, the separation of prices between some constrained zones is "inactive" whenever there is significant market power.

> "The absence of effective competition within the Inactive Zones gives market power to the few owners of generating units within these Zones that could theoretically drive the corresponding zonal price arbitrarily high should there be Congestion Management on the inactive interface."[44]

The limitation is clear, and the appeal to the perverse effects of local market power is taken for granted. The underlying arguments are seldom exposed. It would seem natural that the motivation should be to avoid subsidies for local monopolists and any associated market inefficiencies.

> "When a transmission path within a zone is regularly congested, there are good reasons to create a new zone. This transforms an zonal interface into an inter-zonal interface: doing so allows congestion over this interface to be handled using market processes based on adjustment bids (as opposed to the ISO's procedures for handling zonal congestion). However, creating a new zone can lead to highly

---

[44] California Independent System Operator, "Report to the Federal Energy Regulatory Commission: Studies Conducted Pursuant to the October 30, 1997 Order," December 1, 1999, p. 17.

> concentrated ownership of generation units in the markets within each zone, which enhances the ability of generation units owners to set high zonal prices. In light of this tradeoff, it is not the case that creating more zones always or typically enhances the efficiency of the overall market. Recognizing this tradeoff, the ISO Tariff calls for the creation of new zones only if generation markets on both sides of the interface in question are 'workably competitive.'"[45]

It would be useful to examine the costs of this cost allocation policy in terms of the incentive effects and the subsidy to local monopolists. A separate section considers the impact of zonal price aggregation under competitive market assumptions, but here the issue is market power mitigation. To what extent is it true that modification of the market design to create zonal prices mitigates market power?

It is important to make a distinction between real expansion of the transmission network and zonal aggregation. Real expansion of the network removes transmission constraints and expands competition. By contrast, administrative aggregation into a zone does nothing by itself to eliminate the transmission constraints. The incentive and market effects of real grid expansion are quite different from those of administrative cost reallocation. In fact, for purposes of market power mitigation locational pricing can be superior to zonal aggregation and pricing. There are at least four reasons why locational pricing is often superior to zonal pricing from a when there is potential for the exercise of locational market power.[46]

- Zonal pricing can create market power in the hypothetical zonal dispatch that does not exist in the actual power market under either nodal or inter-zonal pricing.
- Zonal pricing can create market power in the zonal redispatch that does not exist in the actual power market under either nodal or inter-zonal pricing.
- By reducing the response of demand in the constrained region to the exercise of locational market power, zonal pricing can make profitable the exercise of market power that would be unprofitable under either nodal or inter-zonal pricing.
- The zonal pricing and redispatch mechanism can reduce the supply elasticity of energy across open interfaces, making profitable the exercise of market power that would be unprofitable under nodal pricing.

Despite the conventional wisdom, therefore, locational pricing is generally superior for both reasons of static and dynamic efficiency. The conventional wisdom that aggregation into zones would mitigate market power mistakes real transmission expansion for administrative transmission expansion. With real constraints in the grid, administrative aggregation actually exacerbates the market power problem.

Of course, locational pricing by itself does not eliminate market power in the generation market, and other mitigation measures are needed. There are other tools available that can be made

---

[45] Frank A. Wolak, Chairman, Market Surveillance Committee of California Independent System Operator, "Report of Redesign of California Real-Time Energy and Ancillary Services Markets," October 18, 1999, p. 16.

[46] Scott M. Harvey and William W. Hogan, "Imperfect Pricing for Imperfect Markets: Nodal and Zonal Congestion Management and the Exercise of Market Power," Harvard-Japan Project on Energy and the Environment, Harvard University, March 2000.

compatible with the operation of a competitive market following the outline of the essential elements in the framework of the RTO Order. A less than exhaustive list would include:

- Divestiture of generating plants. This is the most obvious first step. Diversification of ownership of generating plants reduces market power. The lesson the of experience around the world is that it is easy to underestimate the number of different owners required to substantially eliminate market power, and position of the plants in the merit order is important.

- Expansion of the transmission system. Real expansion of the transmission system tends to reduce constraints and increase the real size of the market. This would be true for both merchant investment in transmission and regulated investments. In the case of either, there is an argument to be made for providing liberal incentives, or even subsidies for transmission expansion, on the grounds that transmission is cheap and market power is expensive.

- Bid caps rather than price caps, with the bids helping to determine market prices. The application of price caps tends to apply to everyone, even those who do not have market power. Further, price caps confound the problem of distinguishing between high prices due to monopoly rents (bad incentives) and high prices due to scarcity rents (good incentives). By contrast, bid caps apply only to those generators that have market power, can be tailored to their particular circumstances, and make it possible to support the competitive structure and preserve the incentives of scarcity rents. The bid cap, with the associated requirement to offer its capacity, forces the generator with market power to act as though it were a competitive participant. This entitles it and everyone else to receive the market-clearing price, which would include scarcity rents but not monopoly rents.

- Locational contracts for differences for delivery of energy and services. Selling energy forward through a long-term contract fundamentally changes the incentives of the generator, in effect moving form being a large seller in the spot market to being a small seller or even a net buyer. Once under the contract (which must be imposed by a regulator), the generator's interest conforms to the assumptions of the workably competitive market.

- Transmission congestion contracts. Selling congestion relief forward provides a complementary contract for the locational contract for differences. Here the contract is for the difference in price between locations rather than the price at a location. Again, once under the contract (which must be imposed by a regulator), the generator's interest conforms to the assumptions of the workably competitive market.

- Entry of new generators at locations that reduce market power. The fundamental dynamic of long-run competition operates through entry or threat of entry by competitors. Hence it is important to avoid any artificial barriers to entry. Furthermore, the logic of dynamic efficiency says that new entrants tend to reduce market power (at least as long as they don't exploit transmission bottlenecks[47]). Therefore, new entrants

---

[47] Judith B. Cardell, Carrie Cullen Hitt, William W. Hogan, "Market Power and Strategic Interaction in Electricity Networks," Resource and Energy Economics, 19(1997) 109-137.

who do not control some essential facility, should be allowed to operate without bid caps or other market power mitigation limitations.

- Demand side bidding. One of the most powerful and ubiquitous tools for limiting market power is the response of customers who refuse to buy at high prices. Present retail pricing tariffs tend to socialize wholesale price movements, through rate freezes or zonal averaging. The sooner that customers have a realistic opportunity and incentive to bid their demand, rather than accept whatever the market will bear in the aggregate, there will be an important tool available for mitigating market power in the most natural way.[48] To be sure, this would not eliminate market power, but it would at least make the conditions no worse in electricity than they are in other markets where customers pay the market price.

- Regional coordination to eliminate barriers to entry and trade. The arguments above about zonal aggregation and administrative limitations on markets apply to the trade across regional boundaries. To the extent that pricing and other seams issues can expand and integrate regional markets, they will have a general tendency to mitigate market power by, in effect, increasing transmission capacity.

A common theme of all these measures is that they make market power mitigation work without requiring a redesign of the competitive market structure. The measures are either targeted to mitigating market power within the competitive framework, as with bid caps, or changing the decisions of regulated entities, as with transmission expansion.

LESSONS FOR RTOS

There are a few essential services related to coordinating use of the transmission grid where the RTO is both necessary and would have a significant comparative advantage. The RTO would have a significant advantage in conducting its short-term coordination activities through an open spot market. In addition to the immediate efficiency improvements, the transparency and ease of entry for small participants would provide a wealth of benefits for promoting the long-run competitiveness of the market. These benefits would include a practical framework for implementing transmission rights, as embodied in the PJM approach.[49]

The PJM model and similar systems such as for New York, New England, New Zealand, and so on, provide open access with non-discriminatory pricing. The critical short-run matter of congestion pricing and allocation of scarce transmission capacity through locational marginal cost pricing complements other components to deal with the longer-term issues that go beyond system operations. Transmission system fixed costs are recovered primarily through system-wide (but not necessarily uniform) network service charges. The ISO administers both a spot market and bilateral schedules, while maintaining reliability under principles of bid-based, security-constrained economic dispatch. Fixed transmission rights (FTRs, i.e., transmission congestion contracts) are

---

[48] Douglas Caves, Kelly Eakin, and Ahmad Faruqui, "Mitigating Price Spikes in Wholesale Markets through Market-Based Pricing in Retail Markets," Electricity Journal, Vol. 12, No. 3, April 2000, pp. 13-23.

[49] 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).

available for congestion costs between locations, creating the market equivalent of perfectly tradable physical transmission rights, providing tradable open-access transmission reservations in the only way that is likely to be feasible.

There is little public policy justification for approving RTO rules that go against this model and in the direction of adding further complications while restricting participant choices. Hence, it would be appropriate to prescribe rules for the RTO that would support a competitive market in the real world, rather than an idealized world where network complications could be ignored. At a minimum, governments and regulators interested in market design should look with great skepticism on proposals that begrudgingly acknowledge that a certain function must be performed by the system operator, but then require that it be performed badly. The effect of proposals that short-circuit efficient coordination of the spot market is to increase costs, create subsidies and reduce market flexibility.

Many lessons can be drawn from the developing experience with electricity markets. Here there are complementary lessons to be gleaned from earlier work about both what to avoid and what to emphasize. First the pitfalls:[50]

- **Impose Balancing Penalties.** The RTO must provide real time balancing to maintain system integrity. Balancing imposes costs, and those relying on the balancing services should pay these costs. However, a strong burden of proof should face those who would charge balancing penalties in excess of costs, or restrict voluntary access to balancing services.
- **Require Individual Balancing Constraints.** The RTO must maintain aggregate energy balance in the system, but there is no physical necessity and no public policy interest in requiring particular combinations of individual transactions to remain balanced. Quite the contrary. Individual balancing requirements both complicate the task for the RTO and provide a device to reinforce market power. This goes against the public interest.
- **Prohibit Least-Cost (Re)Dispatch.** The RTO must be able to (re)dispatch plants in order to manage transmission congestion. Rules designed to prevent the RTO from applying the familiar principles of economic dispatch run contrary to the notion of competitive markets and the public interest.
- **Reject Voluntary Bids.** When doing an economic dispatch, it seems logical for the RTO to make the adjustments taking into account the preferences of the market participants as expressed by their voluntary bids. There should be a strong burden of proof for those who argue that it is necessary to restrict the voluntary bids, or discard consideration of some bids.
- **Separate Transmission Rights and Dispatch.** The RTO must coordinate the use of the transmission system. And once the actual use of the transmission system is determined, so is the dispatch. Regulators should look with skepticism on any proposal built on the flawed argument that transmission usage and dispatch can be separated.
- **Restrict the Capacity of the Grid.** The real reliability conditions for the electric grid include an ensemble of contingency conditions and complicated network interactions. Relatively few of

[50] This summary comes from William W. Hogan, "Restructuring the Electricity market: Institutions for Network Systems," Harvard-Japan Project on Energy and the Environment, Center for Business and Government, Harvard University, April 1999, pp. 14-16.

these real constraints are simple limits on the actual flow across certain interfaces. Regulators should look skeptically at proposals that require derating the real capacity of the grid in order to make a few flow limits sufficient to guarantee reliability under a simple market model.

- **Suppress Pricing Information.** Only the RTO would have the information needed to calculate and post locational prices, as in PJM. The computations are easy for a given dispatch, but only the RTO has all the information about the dispatch. Given the striking gap between the previous claims that congestion is insignificant and the observed reality of true locational pricing in the first real implementation in the United States, regulators everywhere should have a strong interest in prescribing that the real locational marginal costs--considering the real network interactions, and not just simplified zonal aggregations--be made available on a regular basis.

This recital of illustrative pitfalls was motivated by flawed components found in quite real proposals, a few already in place. Some market participants may prefer large transaction costs, trading obscurity, barriers to entry, and the ability to exploit market power. They should oppose an efficient spot market coordinated by the RTO, which would simplify the real operations of markets and reduce the profits of those who otherwise would benefit from the inefficiencies. An open spot market coordinated by the system operator would be easy to use, and easy access is not profitable if you are selling access. But we should not confuse the public interest in greater competition with an interest in greater profits for ever more competitors. Governments and regulators imbued with responsibility for the public interest should prevent such mistakes.[51]

A set of similar issues arises in the extensive summary of positive lessons learned as reported in the RTO Order. These lessons reflect the early experience with independent system operators and provide another perspective on the implications for electricity market design. An excerpt of the longer discussion in the RTO Order includes:[52]

"We expect that bid-based markets will be a central feature in many RTO proposals. To date, the Commission has analyzed and approved, with various modifications, bid-based market designs for four ISOs. The purpose of this section is to summarize the lessons learned from these real-world market experiments.

a. **Multiple Product Markets**: ... if more than one product is being sold in the same temporal market, efficiency is maximized when arbitrage opportunities reflected in the bids are exhausted (i.e., after the RTO's markets have cleared, no technically qualified market participant would have preferred to be in another of the RTO's markets). In addition, efficient bid-based markets elicit prices that are consistent with technical and cost requirements. ...

b. **Physical Feasibility**: Proper design of the market clearing procedures ensures that prices balance the supply and demand for energy, and all transactions, in the aggregate, are physically feasible with appropriate levels of reserves. ...

---

51 Larry Ruff, "Competitive Electricity markets: Why They Are Working And How To Improve Them," National Economic Research Associates, May 12, 1999.

52 RTO Order, pp. 632-643.

c. **Access to Real-Time Balancing Market**: Making a real-time balancing market available to all grid users ensures that all users are treated equally for purposes of settling their individual imbalances. ...

d. **Market Participation:** Markets are most efficient when generators and loads, whether internal or external to the RTO, are allowed full and flexible participation in the markets. While generators and loads have the option to choose between participating in any RTO-facilitated markets or other markets, the RTO must have generation and ancillary service quantity information, and any necessary technical information, from self-schedulers in order to balance the system and ensure reliability. This allows bilateral and forward financial markets and independent PX markets to co-exist and complement RTO physical markets. Participants that self-schedule would be expected to pay for the costs that they impose on the physical system at market prices and be paid for the benefits that they supply to the physical system at market prices. ...

e. **Demand-Side Bidding**: Demand-side bidding is desirable to the extent it is technically feasible, because without it, demand response decreases and market power is easier to exercise. The availability of price responsive demand also reduces price volatility in the markets.

f. **Bidding Rules**: A market that provides the flexibility for all generators to bid a reasonable approximation of the costs they incur including start-up, minimum load, energy, and ramping costs will be efficient. ...

g. **Transaction Costs and Risk**: Transaction costs associated with participation in well functioning RTO markets should be low, and market participation should involve no unnecessary risks. ...

h. **Price Recalculations**: In circumstances where time does not permit all changes in dispatch to be communicated and effected through manual processes in a timely manner, the market clearing price resulting from the computer algorithm must be adjusted to reflect the actual dispatch in the hour. ...

i. **Multi-Settlement Markets**: Multi-settlement markets may involve a day-ahead and real-time market. For real-time markets, prices are determined by real-time dispatch quantities, and deviations from day-ahead schedules are priced at the real-time price. When day-ahead schedules are financially binding, they are financial commitments subject to payments for deviations at the real-time price. If market participants adhere to day-ahead schedules, they need not participate in the real-time markets. If needed for reliability, bids need to be physically binding and may be subject to Commission-approved penalties for failure to adhere to the bid. Without financially binding commitments in the day-ahead market, the riskiness of market participation increases since the day-ahead bids could be changed before real-time dispatch. ...

j. **Preventing Abusive Market Power**: An efficient market design does not favor market participants that have the potential to exercise market power and minimizes the incentives for market participants to engage in abuse of market power. For example, since large players are more likely to cause market power problems, a market design that favors large players (e.g., portfolio bidding) may create an incentive for consolidation and resulting market power problems. ...

k. **Market Information and Market Monitoring**: One property of an efficient market has market clearing prices and quantities being made available immediately. This information enables market participants and potential future market participants to assess the market and plan their businesses efficiently. It will also allow market participants to spot errors in the market clearing process and get them corrected. ...

l. **Prices and Cost Averaging**: Market designs that base prices on the averaging or socialization of costs, may distort consumption, production, and investment decisions and ultimately lead to economically inefficient outcomes. Where possible and cost effective, cost causality principles can be used to price services and eliminate averaging. Moreover, if pass-throughs or uplift charges are paid by all load to ensure bid-cost recovery, as in some approved ISO market designs, it may be appropriate to couple these pricing mechanisms with incentive mechanisms for the RTO to control them."

The cumulative effect of these negative and affirmative lessons emphasizes the importance of a close integration of the market design with the technical characteristic of electric network systems. A fundamental feature of competition is to give the market participants choice. This means that the incentives and prices guiding choices must be carefully matched with the reality of the operation of the system. Attempts to subvert either the market or the physics inevitably produce problems in the form of unintended consequences.

ALTERNATIVE MARKET MODELS

Electricity systems are not simple. The reality of electricity systems and the special problems induced by network interactions create an interest in simplifying market design to provide better support for commercial transactions. The benefits of simplification are clear, and other things being equal it would be desirable to adopt a simpler design. However, other things are usually not equal, and the law of unintended consequences often dictates that what appears simple may turn out to be complex in the end. Further, what may appear complex on the surface turns out to be simple in the end because it is consistent with the reality of the electric system and does not require substantial non-market interventions to make the market work.

The debate over alternative electricity market institutions often confuses two design issues that could, in principle, be treated separately. The distinction is between what is appropriate as a basis for the design of an RTO, and what would be appropriate as the design of a stand alone business offering a service within the framework of an RTO. As illustrated by the discussions below, the usual rehearsal of the confusion begins with an assertion that some element of the market framework could be radically simplified without serious harm to market operations and with substantial benefits for commercial transactions. The system operator and others demur on the grounds that the simplifications suggested are overly simplistic; the proposed approximations would create both severe operational problems and produce bad incentives that could be exploited by market participants. By contrast, proponents argue that the approximations required are not commercially significant and that the market participants would actually prefer to take advantage of the simpler model.

It could be that both sides are right. It is one thing for the RTO to deviate from the dictates of reliability or economic dispatch, using its monopoly power and ability to involuntarily socialize or shift costs among the market participants. It is quite another matter for a private business to voluntarily take on the risks of the market simplifications and provide a service to other market participants. Hence there is a simple market-oriented resolution of what is often a contentious debate. The RTO, in its role as supporting the competitive market, should design the rules and determine the prices in ways that are faithful to the reality of the electric system. If the approximations and simplifications preferred by some in the market truly have little or no commercial impact, then these risks could be absorbed by a private business and should not be the responsibility of the RTO. The market participants could offer services that repackage and bundle the basic products maintained by the RTO, and the RTO could be careful to offer the essential market ingredients as outlined above. In the end, this will be the truly simple market design.

Two such debates, reappearing in virtually every design or reform effort for an electricity market, illustrate and reinforce this basic analysis. Here we consider the arguments for zonal price aggregation and then for decentralized congestion management.

Congestion Zones

Full locational pricing at every node in the network is a natural consequence of the basic economics of a competitive electricity market. However, it has been common around the world to assert, usually without apparent need for much further justification, that nodal pricing would be too complicated and aggregation into single price zones, with socialization of the attendant costs, would be simpler and solve all manner of problems. On first impression, the argument appears correct. On closer examination, however, we find the opposite to be true, once we consider the incentives created by aggregation combined with the flexibility allowed by market choices. The debate continues, but the negative evidence is accumulating.

For example, the first region in the United States to abandon a zonal pricing model after it failed in practice was PJM, from its experience in 1997 when its zonal pricing system prompted actions which caused severe reliability problems. Given this experience, PJM adopted a nodal pricing system that has worked well since March 1998.[53] Subsequently, the original one-zone congestion pricing system adopted for the New England independent system operator (ISONE) created inefficient incentives for locating new generation.[54] To counter these price incentives, New England proposed a number of limitations and conditions on new generation construction. Following the Commission's rejection of the resulting barriers to entry for new generation in New England, there developed a debate over the preferred model for managing and pricing transmission congestion.[55] One zone was not enough, but perhaps a few would do? In the end, New England

---

[53] William W. Hogan, "Restructuring the Electricity market: Institutions for Network Systems," Harvard-Japan Project on Energy and the Environment, Center for Business and Government, Harvard University, April 1999, pp. 37-44.

[54] The use of zones for collecting transmission fixed charges is not the issue here. The focus is on managing transmission congestion. 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.

[55] Federal Energy Regulatory Commission, New England Power Pool Ruling, Docket No. ER98-3853-000,

proposed go all the way to a nodal pricing system.[56]

A similar zonal congestion management market design created similar problems in California, which prompted the Commission to reject a number of 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."[57] As a further example, the zonal pricing system in Alberta, Canada, apparently produced a related set of incentives that failed to give generators the price signal to locate consistent with the needs of reliability: "Most of the electricity generation sources are located in the northern part of the province and ever-increasing amounts of electricity are being transported to southern Alberta to meet growth, … [t]his is causing a constraint in getting electricity into southern Alberta and impacting overall security of the high-voltage transmission system."[58] As a result, Alberta has proposed a central generation procurement process under the transmission operator to provide a means to get generation built in the right place. Hence we have the ironic result of a supposed simplification of the market under zonal pricing that seems headed towards replacing central procurement by the monopoly utility with central procurement by the monopoly transmission provider. This is hardly a true simplification, nor is it consistent with the original intent to move towards a competitive market and away from monopoly procurement.

> **Fact:** A single transmission constraint in an electric network can produce different prices at every node. Simply put, the different nodal prices arise because every location has a different effect on the constraint. This feature of electric networks is caused by the physics of parallel flows. Unfortunately, if you are not an electrical engineer, you probably have very bad intuition about the implications of this fact. You are not alone.
>
> **Fiction:** We could avoid the complications of dealing directly with nodal pricing by aggregating nodes with similar prices into a few zones. The result would provide a foundation for a simpler competitive mark**e**t structure.

The accumulating evidence reveals the flaws in this seductive simplification argument.[59] In reality, the simplification creates unexpected problems. These problems in turn cause the system operator to intervene in the market by imposing non-market solutions and socializing the costs. In the end, the truly simple system turns out to be a market that uses nodal pricing in conjunction with

---

October 29, 1998.

56 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.

57 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.

58 "Alberta Transmission Czar Wants More Generation," Electricity Daily, Vol. 14, No. 77, April 21, 2000, p. 3.

59 William W. Hogan, "Nodes and Zones in Electricity Markets: Seeking Simplified Congestion Pricing" in Hung-po Chao and Hilliard G. Huntington (eds.), Defining Competitive Electricity Markets, Kluwer Academic Publishers, 1998, pp. 33-62. Steve Stoft, "Transmission Pricing in Zones: Simple or Complex?", The Electricity Journal, Vol. 10, No. 1, January/February 1997, pp. 24-31.

a bid-based, security-constrained, economic dispatch for the real network, administered by an independent system operator. Purchases and sales in the balancing spot market would be at the nodal prices. Bilateral transactions would be charged for transmission congestion at the difference in the nodal prices at source and destination. Transmission congestion contracts would provide price certainty for those who pay in advance for these financial "firm" transmission rights up to the capacity of the grid. The system would be efficient and internally consistent.

Note that the problem with zonal price aggregation and poor incentives does not extend to the use of market hubs within the framework of nodal pricing.[60] The hub-and-spoke model fits quite naturally within the nodal pricing framework and has been operating successfully in PJM, producing a liquid forward market at the PJM "Western Hub."[61] Market hubs can and do provide virtually all the benefits of simplification often attributed to zonal price aggregation. The difference between a hub-and-spoke model and zonal price aggregation is simple: zonal aggregation gives you the hubs without the spokes. The spokes capture the difference between the nodal price and the market hub price. Zonal price aggregation assumes these differences can be ignored, and then socializes the cost when than cannot. And just as the wheel would not support the hub without the spokes, the missing spokes in the zonal model lead to a collapse of price incentives followed by the inevitable requirement for operator intervention.

In some cases, of course, the arguments offered for zonal price aggregation may be true. The differences in nodal prices may be small, most of the time, and the occasional excursions would not be commercially significant. Or, to be more precise, the occasional excursions would not be significant as long as the system operator did not socialize the costs. Under these circumstances, there is a clear business opportunity. The RTO need not and should not do anything different. Within this framework, an entrepreneur would be free and able to set up a business that provided the aggregation service, charging participants for the claimed benefits and providing a revenue stream to compensate for the small risks involved.

When viewed from this perspective, the arguments in favor of zonal price aggregation should not be seen as applying to the RTO. As we have learned, when the RTO follows this path, trouble soon appears. Rather, the arguments for zonal aggregation should be seen as either wrong or right. If wrong, they should be ignored. If right, they should lead to a successful business. But zonal price aggregation is usually a bad market design for an RTO.

Flowgates and Decentralized Congestion Management

The essential market ingredients outlined above include a coordinated spot market integrated with system operations to provide balancing services and congestion management. In principle, an alternative to central coordination would be a system of decentralized congestion management that used the same basic information as the system operator but could be handled directly by the market participants.

---

60 William W. Hogan, "Restructuring the Electricity Market: Institutions for Network Systems," Center for Business and Government, Harvard University, April 1999, p. 52, available from the author's web page.

61 "The New York Mercantile Exchange will launch an electricity futures contract March 19 at the PJM western hub, one of the most liquid markets in the Eastern grid. ... The PJM hub already features an active and growing over-the-counter forwards market. A liquid hub can have a downside [for the futures contract] given that players are content trading in the OTC, said one Northeast broker." Power Markets Week, February 8, 1999, p. 14.

The most prominent example of such a decentralized congestion management model is the so-called "flowgate" approach. This is interesting as both a theoretical argument[62] and because it is the procedure embraced by NERC as a principal market alternative to its administrative TLR procedures.[63] The details can be complicated, but the basic idea is simple. The argument begins with the recognition that the contract path model is flawed. Power does not flow over a single path from source to sink, and it is this fact that causes the problems that lead to the need for TLR in the first place. If a single contract path is not good enough, perhaps many paths would be better. Since power flows along many parallel paths, there is a natural inclination to develop a new approach to transmission services that would identify the key links or "flowgates" over which the power may actually flow, and to define transmission rights according to the capacities at these flowgates. This is a tempting idea with analogies in markets for other commodities and echoes in the many efforts in the electricity industry for MW-mile proposals, the General Agreement on Parallel Paths (GAPP), and related efforts that could go under the heading of transmission services built on link-based rights.

For any given total set of power injections and withdrawals, it is possible to compute the total flows across each line in the transmission network. Under certain simplifying assumptions, it would be possible further to decompose the flows on the lines and allocate an appropriate share of the flows to individual transactions that make up the total loads. If we also knew the capacity on each line, then presumably it would be possible to match the flows against the capacities and define transmission services. Transmission users would be expected to obtain rights to use the individual lines, perhaps from the transmission line owner or from others who owned these capacity rights.

In principle, these rights on each line might be seen as supporting a decentralized market. Associated with each line would be a set of capacity allocations to (many) capacity right holders who trade with the (many) users of the system who must match their allocated flows with corresponding physical capacity rights. Within this framework there are at least two interesting objectives. First, that the trading rules should lead to an efficient market equilibrium for a short period; and second, that the allocated transmission capacity rights would be useful for supporting the competitive market for geographically dispersed buyers and sellers of power.

As a matter of principle, it is likely that the first objective could be met. There should be some system of tradable property rights that would be sought by users of the system, and in so doing would lead to an efficient short-run dispatch of the system. This would seem to be nothing more than an application of the principles of competitive markets with well-defined property rights and low transactions costs. There is a general belief that this short-run efficiency would be available in principle: "Efficient short-run prices are consistent with economic dispatch, and, in principle, short-run equilibrium in a competitive market would reproduce both these prices and the associated power flows."[64] The problem has always been with the natural definitions of the "physical" rights: these are cumbersome to trade and enforce. The property rights are hard to define, and the transaction

---

[62] Hung Po Chao and Stephen Peck, "A Market Mechanism for Electric Power Transmission," Journal of Regulatory Economics, Vol. 10, No. 1, 1996, pp. 25-59. Steven Stoft, "Congestion Pricing with Fewer Prices than Zones," Electricity Journal, Vol. 11, No. 4, May 1998, pp. 23-31.

[63] Congestion Management Working Group of the NERC Market Interface Committee, "Comparison of System Redispatch Methods for Congestion Management," September 1999.

[64] W. Hogan, Contract Networks for Electric Power Transmission," Journal of Regulatory Economics, Vol. 4, 1992, p. 214.

costs of trading would not be low.

The second objective is perhaps more important. Presumably the allocated transmission capacity rights would extend over many short-run periods, for example, even only a few days, weeks or months of hourly dispatch periods.[65] Presumably a natural characteristic that would be expected of these physical rights would be that a seller of power with a known cost of power production could enter into an agreement with a distant buyer to deliver a known quantity of power at a fixed price, including the out-of-pocket cost for transmission using the transmission right. Many other contracts could be envisioned, but this minimal possibility would seem to be essential; and it is broadly taken for granted that this capability will exist in the open-access transmission regime. However, any approach that defines tradable physical capacity rights based on flows on individual lines faces obstacles that appear to make it impossible to meet this minimal test.

There are many variants of such link-based transmission rights that one can imagine, and the industry has been struggling with these ideas for years. Here the flowgate argument follows the outline above. The system operators and others demur on the grounds that the electric system is more complicated and there are simply too many lines and possible constraints to manage in a decentralized environment. The proponents argue that it is not necessary to consider all the lines and all the possible constraints. Rather they propose to consider only a few critical constraints, the flowgates, and to focus decentralized trading on these. The assertion is that the commercially significant congestion can be represented by a system with:

- Few flowgates or constraints.
- Known capacity limits at the flowgates.
- Known power transfer distribution factors (PTDF) that decompose a transaction into the flows over the flowgates.

Under these simplifying assumptions, the decentralized model might work in practice. The RTO would identify the flowgates. The capacity rights would be allocated or auctioned somehow to the market participants. Similarly, the RTO would publish the PTDF table that would allow individual market participants to compute the effect of their transactions on the flowgates. The participants would then purchase the corresponding flowgate capacity rights in the market. This trading of capacity rights would take place in decentralized forward markets. Transactions that had assembled all the capacity rights needed would then be scheduled without further congestion charges. Real-time operations would be handled somehow, typically not specified as part of the flowgate model.

There is some experience with this flowgate model. However, the experience is limited and what experience we do have is not good. In particular, these simplifying assumptions and the corresponding flowgate model for decentralized congestion management were applied as part of the NERC Pilot Project for Market Redispatch in 1999, to create a decentralized alternative to administrative TLR curtailments. In the end, and despite the substantial turmoil created by the TLR system, the result was that apparently there were *no* successful applications of any decentralized trades under this approach.[66] By contrast and at the same time, the centralized coordinated market in

---

65 This is apart from the problems encountered with changes of the grid capacity or configuration. Link-based rights have other substantial problems for dealing with system expansion.

66 Congestion Management Working Group of the NERC Market Interface Committee, "Final Report on the

PJM regularly provided successful market alternatives to administrative TLR curtailments. Perhaps the flowgate problems will be ironed out as the NERC experiment continues,[67] but the experience reinforces the need to look more closely at the flowgate model.

Despite the appeal of a move away from the contract path model and closer to the actual underlying reality of the transmission network, these generic methods built on flowgate rights must confront the problems inherent in the simplifications. Are there only a few flowgates? Are the capacity limits known in advance? Are the PTDF impacts stable and known in advance of real-time?

Those who demur in accepting the flowgate model as a method for organizing the use of the transmission system would answer in the negative for each of these three questions. First, there are many potential constraints, so it would be necessary to obtain many capacity rights on flowgates. The number of rights that would have to be acquired in a complete version of a flowgate model generally would not be determined simply by the amount of power that flows in the actual dispatch. Under current practice, the system operators typically adhere to "(n-1) contingency" constraints on power flows through the grid. This means that the allowed power loads at every location in the transmission system must be such that in the event one of series of possible contingencies occurs, the instantaneous redistribution of the power flows that results will still meet minimum standards for thermal limits on lines and will still avoid voltage collapse throughout the system. We can think of the terminology as coming from the notion that one of the "n" lines in the system may drop out of service, and the system must still work with the (n-1) lines remaining. The actual contingencies monitored can be more diverse, but this interpretation conveys the basic idea of an (n-1) contingency-constrained power flow.

Hence, a single line may have a normal limit of 100 MW and an emergency limit of 115 MW.[68] The actual flow on the line at a particular moment might be only 90 MW, and the corresponding dispatch might appear to be unconstrained. However, this dispatch may actually be constrained because of the need to protect against a contingency. For example, the binding contingency might be the loss of some other line. In the event of the contingency, the flows for the current pattern of generation and load would redistribute instantly to cause 115 MW to flow on the line in question, hitting the emergency limit. No more power could be dispatched than for the 90 MW flow without potentially violating this emergency limit. The 90 MW flow, therefore, is constrained by the dispatch rules in anticipation of the contingency. The corresponding prices would reflect these contingency constraints.[69]

Depending on conditions, any one of many possible contingencies could determine the current limits on the transmission system. During any given hour, therefore, the actual flow may be, and often is, limited by the impacts that would occur in the event that the contingency came to pass. Hence, the contingencies don't just limit the system when they occur; they are anticipated and can limit the system all the time. In other words, analysis of the power flows

---

NERC Market Redispatch Pilot," November 29, 1999, filed with FERC on December 1, 1999.

67 NERC, "Market Redispatch Pilot Project Summer 2000 Procedure," March 31, 2000.

68 Expressing the limits in terms of MW and real power is shorthand for ease of explanation. Thermal limits are actually in terms of MVA for real and reactive power.

69 Jacqueline Boucher, Benoit Ghilain, and Yves Smeers, "Security-Constrained Dispatch Gives Financially and Economically Significant Nodal Prices," Electricity Journal, November 1998, pp. 53-59.

during contingencies is not just an exception to the rule; it is the rule. The binding constraints on transmission generally are on the level of flows or voltage in post-contingency conditions, and flows in the actual dispatch are limited to ensure that the system could sustain a contingency. Operation of a complete flowgate model, therefore, would require a trader to acquire the rights on each link sufficient to cover its flows on that line in each post-contingency situation.

A sometime argument is that this problem is not serious because the actual dispatch will have only a few of the potential constraints actually binding. Typically this is true, but it does not avoid the difficulty for the simple reason that we don't know in advance which constraints will be binding. Were it otherwise the system operator would not have to monitor all the constraints that are typically considered. In fact, the large list of potential constraints monitored by the system operator is already a select group identified as the important subset from the thousands or millions of possible constraints that could be defined given the large number of lines and the large number of contingencies. The mere fact that the system operator as identified the constraints would arguably be enough to require an associated flowgate capacity right in order to ensure that the resulting transaction would be feasible.

The accumulating experience in PJM is well documented and amply illustrates the point. In one outside study intended to support the development of a zonal model and decentralized congestion management through something like a flowgate model, a set of 28 constraints were identified as important and analyzed for the variations in the equivalent of a PTDF table. While 28 may seem a large number and difficult to deal with in assembling the capacity rights to use the transmission system, it turned out not to be large enough. In the event, the first six months of operation of locational pricing in PJM found 43 constraints actually binding. Most importantly, *none* of these actual constraints were in the list of 28 supposedly easy-to-identify flowgates.[70] This suggests the magnitude of the difficulties faced when predicting which constraints will be binding. And the list of real constraints continues to grow. Over the period January 1998 to April 2000, there were 171 unique constraints that produced congestion and different locational prices in PJM.[71] Apparently a complete flowgate model would require purchase of at least 171 capacity rights to secure a single point-to-point transaction.

The obstacle of too many constraints to specify a complete flowgate model might be overcome if it were still possible to identify in advance how much capacity there is at each flowgate. This is an old problem with the uncomfortable reality that for many of the constraints it is not possible to specify the limiting value without also knowing the pattern of the loads. For example, interface constraints for voltage protection are routinely described as a range of maximum values on real power flows, with the actual value being set and changed regularly during real time operations. The PJM Eastern Reactive Transfer Limit is reset at least every 15 minutes and can vary over a range of 4000 MW to 7000 MW, depending on system conditions.[72] This is essentially the same

---

70 Richard D. Tabors, "Transmission Pricing in PJM: Allowing the Economics of the Market to Work," Tabors Caramanis & Associates, February 24, 1999, p. 31. This is a careful study that is among the rare instances with easily available and documented assumptions. See the PJM web page for the record of actual constraints.

71 See the PJM web page spreadsheet report on historical transmission limits, "Historical_TX_Constraints.xls." Over the period January 1998 to April 2000 there were 610 constraint-days recorded, with the same constraint appearing on more than one day. Based on the "Monitor" and "Contingency" names, there were 171 unique constraints.

72 Andy Ott, PJM, personal communication.

problem as defining the available transmission capacity. As the New York Power Pool (NYPP) observed in a typical comment heard from system operators:

> "The primary responsibility of the NYPP system operator is and must be to maintain the reliability of the bulk power system. The operator must have the flexibility to decide, for example, what level of transmission reserve capacity should be retained under various conditions and facilities' loadings to meet contingencies as they may arise. Thus, actual transmission availability, or, more correctly, available transmission transfer capability, may be less than the thermal limits of the facilities, and the difference may change as conditions change. The Commission should make certain that all participants understand and accept these factors."[73]

In addition to recognizing that the capacity limits are not always known in advance, the other reality is the lack of truly stable and known PTDF tables. The flows over the lines and voltages at the buses will depend on all the other receipts and deliveries on the grid. Thus, the flow over a particular flowgate that can be attributed to a particular transaction will be changing all the time, so it will be difficult to know how much of a flowgate capacity right is required or how much would be used. There are many causes of this ex ante ambiguity in the PTDFs. First, the PTDFs are a function of the entire configuration of the grid. With any line out of service, there are different PTDFs, and the configuration of the grid is changing all the time. Even with the same configurations on the wires, there are many electrical devices, such as phase angle regulators, whose very purpose is to change the apparent impedance of lines as a function of changing loads and, therefore, to change the PTDFs throughout the system. Furthermore, there are inherent nonlinearities in the flows and constraints, especially the ubiquitous so-called “nomogram” constraints that attempt to approximate even more complex interactions in the system. It is simply not true that the real system conforms to the simplified textbook approximation of the pure DC-load model that is useful for illustrating the effects of network flows, but that is at best only a linearized local approximation of the real system that can be used to guide the dispatch. It is for these reasons that PJM updates both the load flow estimate and calculation of its equivalent of PTDF tables every five minutes.[74] In reality, the PTDFs needed for a complete flowgate model would be anything but known in advance.

These criticisms of the flowgate model sit at the foundation of the argument that it would not be an appropriate model for operating the power system. However, as with the arguments above for zonal congestion management, the criticisms may be less applicable to a commercial model that would serve as an entrepreneurial business. Suppose that the criticisms are correct, but the commercial significance is small. The RTO could operate the coordinated dispatch and define financial transmission rights as outlined above for the real system rather than the flowgate approximation. Although it is not possible to identify all the components of the flowgate model in advance, it is possible to determine in advance if a particular load flow would be feasible. Hence, despite the complexity of the grid, a set of simultaneously feasible point-to-point financial

---

[73] Comments of Member Systems of the New York Power Pool, "Request for Comments Regarding Real-Time Docket Information Networks," No. RM95-9-000, Federal Energy Regulatory Commission, July 5, 1995, p. 9-10.

[74] Andy Ott, PJM, personal communication.

transmission rights can be defined. For a given configuration of the grid, the RTO can guarantee the point-to-point FTRs without using its powers to tax the participants and socialize the costs.

Under the simplifying assumptions of the flowgate model, it would be possible to decompose these point-to-point financial transmission rights into their component flowgates, implied flow capacities on flowgates, and the associated PTDFs. If the approximation errors of the flowgate model are not large, then it would be possible for a new business to provide the service of organizing trading of flowgate rights that could be reconfigured to create new FTRs. The differences in flows and capacities might be small, most of the time, and the occasional excursions would not be commercially significant. Or, to be more precise, the occasional excursions would not be significant as long as the system operator did not socialize the costs. Under these circumstances, there is a clear business opportunity. The RTO need not and should not do anything different than outline above as part of the essential market design.[75] Within this framework, an entrepreneur would be free and able to set up a business that provided the flowgate service, charging participants for the claimed benefits and providing a revenue stream to compensate for the small risks involved. In effect, the business could take the financial risk that the reconfigured FTRs might not be feasible in the real network, but if the flowgate assumptions are valid this risk would be small.

When viewed from this perspective, the arguments in favor of the flowgate approach should not be seen as applying to the RTO. When the RTO follows this path, trouble is likely to appear because the real system is more complicated. Rather, the arguments for the flowgate approximation should be seen as either wrong or right. If wrong, they should be ignored. If right, they should lead to a successful business. But the flowgate model is likely to be a problematic market design for an RTO.

FACING THE INEVITABLE

The Commission faces many challenges and will have to set priorities that concentrate the focus on what is really important. The most important choice is in the design of the market institutions for coordination of short-term operations. The good news is that the RTO Order has the analysis right. In the end, the Commission cannot escape the implications of its own analysis.

The Transco diversion will not provide a simple escape from this responsibility. This follows for a number of reasons. First, it is very unlikely that the pure Transco model is viable at all. This goes beyond the conflicts of interest as identified by the FTC. The argument that system operations must be controlled and determined by the owner of the transmission wires does not stand up under even the most minimal investigation. Taken at face value, the pure Transco model would require a single company to acquire ownership of all the transmission wires in a very large region. This is unlikely in the large interconnected grids in the United States. More realistic would be the outcome that a Transco could acquire ownership of only part of the interconnected grid. In this case, we can imagine two outcomes. One would be that the Transco would extend operational control to cover a wider area than its pattern of ownership; but this then would vitiate the original argument that control must follow ownership. Alternatively, control could be limited to follow the balkanized pattern of ownership; but then the necessities of transmission loading relief would create

---

[75] For hedges on the flowgate capacities that were not fully required in the point-to-point financial rights, the RTO could auction of rights to excess congestion revenues from reconfiguration of the FTRs.

an immediate need for external rules dictating the practices of system operations.

This is not to say that an independent transmission company organized as a Transco is impossible. It would be possible to have a de facto ISO within the organizational structure of the Transco, but the access and market rules for this ISO would have to be specified. And the "ring fence" around the embedded ISO would have to be established to avoid affiliate abuse problems. This is, for example, the Transpower model in New Zealand. But what is implausible is that the Commission could avoid the hard work of defining the rules for system operations along the lines set out in the RTO Order. Furthermore, the advantages of an independent transmission company can be pursued through the Gridco model with an accompanying ISO. This approach is already well advanced in the United States and elsewhere. By separating ownership of the wires from control of system operations, it would be easy to accommodate the complex pattern of ownership without foregoing the benefits of regional transmission organizations.

The related and ubiquitous call for "flexibility" will not save the us either. It is all well and good to call for flexibility and experimentation, but some things need consistency. If the Commission does not act to effectively require a consistent design of the institutions for short term coordination of the spot market, along the lines described in the RTO Order, then various market regions are likely to produce different methods that are in conflict. We know what happens when this is allowed, as we have already seen it. The real problems of the interconnected network are not under the Commission's control. Hence, in order to keep the lights on, we have to invent something like the NERC policy on transmission loading relief. Almost immediately, under the name of reliability, the Commission is back in the business of mandating rules for consistent short-term system operations, undoing what has been done in the name of flexibility. However, the Commission would do so through a very limited and stilted conversation that so far has been unable to fully confront the very strong interaction between reliability rules and commercial practices. In the end, the Commission would be in the prescription business; but would not be doing a very good job. Better to face the inevitable and get the market design right in the first place.

The cost of further delay would substantial. The use of the transmission grid involves substantial network externalities through interactions among the schedules of the market participants. If we don't have a pricing system that addresses these interactions, then we must rely on administrative procedures that are inherently discriminatory or subject to easy manipulation. If the prices don't reflect the opportunity costs, not all market participants can be treated in the same way. Some must get preferential grid access. Without an efficient spot market with efficient prices and the associated ease of access, the problems of discrimination will persist.

The problems of transmission congestion cannot be avoided. The legitimate pressure on system operators to avoid reliability problems means that when the market is creating the reliability problem, non-market mechanisms must be employed to undo what the market has done. Without an efficient spot market and its consistent incentives, operational problems will force system operators to impose administrative command-and-control procedures that defeat the purposes of the market participants. The complaints about the serious defects of the competitive market will persist, even grow, as long as the pricing system does not reflect what is really happening.

The development of retail customer access in various regions of the country will place greater pressure on wholesale markets. In particular, we would like to encourage many new entrants and provide balancing and backup services that can support even small players. With a coordinated spot market and readily available efficient prices, this is easy to do. Without an efficient spot market and the associated transparent spot prices, it will be much more expensive and difficult to arrange balancing and settlement for the increasing number of retail access programs in the states.

Furthermore, as we look to the future we see a need for investment in both generation and transmission upgrades. It is possible to exploit market approaches to encourage such investment, and market-based transmission investment is underway in other parts of the world. However, without an efficient spot market and the associated locational prices, there will be no way to define a workable system of transmission rights, no way to stimulate investment in transmission by market participants and, therefore, no way to avoid complete reliance as of old on monopoly decision-making and investment. This will leave the Commission with even more of the burden of defining the incentives and judging the merits of transmission investment, with little or no help from the electricity market.

It is in public interest to improve the design and operation of short-term electricity markets. Once done, many of the other problems in the electric network would either disappear or would be greatly simplified. The problems are real, significant, and here. The Commission must address them, and will, one way or another. The best way to face the inevitable is to recognize it and do the best we can under the circumstances. The Commission knows what to do. Doing it may require using all its powers to persuade, or it may require legislation to clarify its authority to mandate. It may require both.

Apparently the Commission recognizes this pressure. While emphasizing the emphasis on voluntary formation of RTOs, the Commission sets out a timetable:

> "...all public utilities (with the exception of those participating in an approved regional transmission entity that conforms to the Commission's ISO principles) that own, operate or control interstate transmission facilities must file with the Commission by October 15, 2000, a proposal for an RTO with the minimum characteristics and functions to be operational by December 15, 2001, or, alternatively, a description of efforts to participate in an RTO, any existing obstacles to RTO participation, and any plans to work toward RTO participation."[76]

If there are not enough volunteers, then:

> "The goal of this rulemaking is to form RTOs voluntarily and in a timely manner. The alternative to a voluntary process is likely to be a lengthy process that is more likely to result in greater standardization of the Commission's RTO requirements among regions. Although the Commission has specific authorities and responsibilities under the FPA to protect against undue discrimination and remove impediments to wholesale competition, we find it appropriate in this instance to adopt an open collaborative process that relies on voluntary regional participation to design RTOs that can be tailored to specific needs of each region."[77]

In the interest of good public policy and well functioning electricity markets, it would be best to make the voluntary approach work, soon.

---

[76] RTO Order, p. 7.

[77] RTO Order, p. 8.

CONCLUSION

The developing experiences around the world provide insight into the options and implications of alternative models of access to transmission grids in support of an efficient competitive market. It is apparent from this experience that the central 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.

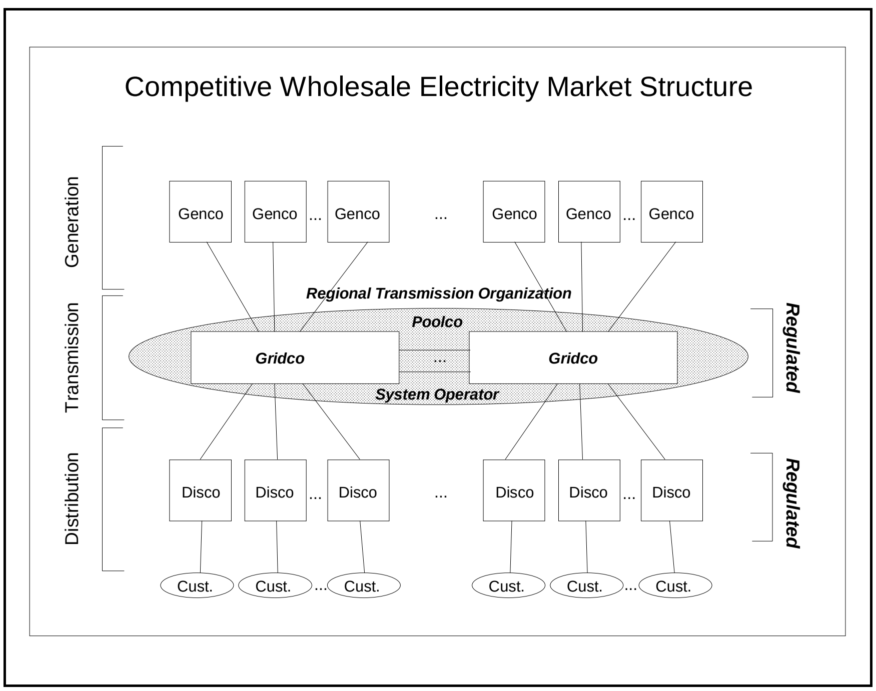
Policy for the continuing evolution of electricity restructuring should emphasize the institutions for market operations. Interconnections through the transmission grid create the necessity for regional organizations that can accommodate competition in services, generation, and contracting while preserving the reliability of the transmission system. Alternative models are many, but can be grouped under the general headings of "Transcos," "Gridcos," "ISO/PX," "ISOs," and finally, organizations for transmission loading relief. The different models present alternatives for the mix of responsibilities. However, the discussion of the differences can distract from recognition of the more important common requirements of any workable system for a competitive electricity market. The key ingredients are known, and they deal with the management of transmission congestion. Putting this system in place should be of the highest priority, as it will simplify many of the other problems facing the Commission. Failing to put this system in place would complicate the development of competitive markets.

APPENDIX
ECONOMICS OF A COMPETITIVE ELECTRICITY MARKET

A general framework that encompassed the essential economics of electricity markets provides a point of reference for evaluating market design elements.[78] Here we focus on the implications for competition in generation, and the rules for the wholesale market. The treatment of competition for other contestable elements, such as retail services, is important but need not affect the design of the wholesale market. This framework provides a background for evaluating the prescriptions for market institutions under the label of Regional Transmission Organizations.

Competitive Market Design

Reliable operation is a central requirement and constraint for any electricity system. Given the strong and complex interactions in electric networks, current technology with a free-flowing transmission grid dictates the need for a system operator that coordinates use of the transmission system. Control of transmission usage means control of dispatch, which is the principal or only means of adjusting the use of the network. Hence, open access to the transmission grid means open access to the dispatch as well. This is the essential coordination function provided by the system operator, no matter what name we give to it. In the analysis of electricity markets, therefore, a key focus is the design of the interaction between transmission and dispatch, both procedures and pricing, to support a competitive market.



To provide an overview of the operation of an efficient, competitive wholesale electricity market, it is natural to distinguish between the short-run operations coordinated by the system operator and long-run decisions that include investment and contracting. Market participants are price takers and include the generators and eligible customers. For this discussion, distributors are included as customers in the wholesale market, operating at arm's length

---

[78] This summary comes primarily from William W. Hogan, "Transmission Investment and Competitive Electricity Markets," Center for Business and Government, Harvard University, April 1998. The issues are developed further there, but summarized here for completeness given the central importance of the basic economics in the case of electricity. See also, William W. Hogan, "Competitive Electricity Markets: A Wholesale Primer," Center for Business and Government, Harvard University, December 1998.

from generators. The system is much simpler in the very short run when it is possible to give meaningful definition to concepts such as opportunity cost. Once the short-run economics are established, the long-run requirements become more transparent. Close attention to the connection between short- and long-run decisions isolates the special features of the electricity market.

Short-Run Market

The short run is a long time on the electrical scale, but short on human scale--say, half an hour. The short-run market is relatively simple. In the short run, locational investment decisions have been made. Power plants, the transmission grid, and distribution lines are all in place. Customers and generators are connected and the work of buyers, sellers, brokers and other service entities is largely complete. The only decisions that remain are for delivery of power, which in the short-run is truly a commodity product.

On the electrical scale, much can happen in half an hour and the services provided by the system include many details of dynamic frequency control and emergency response to contingencies. Due to transaction costs, if nothing else, it would be inefficient to unbundle all of these services, and many are covered as average costs in the overhead of the system. How far unbundling and choice should go is an empirical question. For example, real power should be identified and its marginal cost recognized, but should this extend to reactive power and voltage control as well? Or to spinning reserve required for emergency supplies? For the sake of the present discussion, focus on real power and assume that further unbundling would go beyond the point of diminishing returns in the short-run market.

Over the half hour, the market operates competitively to move real power from generators to customers. Generators have a marginal cost of generating real power from each plant, and customers have different quantities of demand depending on the price at that half hour. The collection of generator costs stacks up to define the generation "merit order," from least to most expensive. This merit order defines the short-run marginal-cost curve for the market, which governs power supply. Similarly, customers have demands that are sensitive to price, and higher prices produce lower demands. Generators and customers do not act unilaterally; they provide information to the dispatcher to be used in a decision process that will determine which plants will run at any given half hour. Power pools provide the model for achieving the most efficient dispatch given the short-run marginal costs of power supply. Although dispatchable demand is not always included, there is nothing conceptually or technically difficult about this extension. The system operator controls operation of the system to achieve the efficient match of supply and demand.

This efficient central dispatch can be made compatible with the market outcome. The fundamental principle is that for the same load, the least-cost or economic dispatch and the competitive-market dispatch are the same. The principal difference between the traditional power pool and the market solution is the price charged to the customer. In the traditional regulated power pool model, customers pay and generators receive average cost, at least on average. Marginal cost implicitly determines the least-cost dispatch, and marginal cost is the standard determinant of competitive market pricing.

An important distinction between the traditional central dispatch and the decentralized market view is found in the source of the marginal-cost information for the generator supply curve. Traditionally the cost data come from engineering estimates of the energy cost of generating power from a given plant at a given time. However, relying on these engineering estimates is problematic in the market model since the true opportunity costs may include other features, such as the different levels of maintenance, that would not be captured in the fuel cost. Replacement of the generator's engineering estimates (that report only incremental fuel cost) with the generator's market bids is the natural alternative. Each bid defines the minimum acceptable price that the generator would accept to run the plant in the given half hour. And these bids serve as the guide for the dispatch.

As long as the generator receives the market clearing price, and there are enough competitors so that each generator assumes that it will not be providing the marginal plant, then the optimal bid for each generator is the true marginal cost: To bid more would only lessen the chance of being dispatched, but not change the price received. To bid less would create the risk of running and being paid less than the cost of generation for that plant. Hence, with enough competitors and no collusion, the short-run economic dispatch market model can elicit bids from buyers and sellers. The system operator can treat these bids as the supply and demand and determine the balance that maximizes benefits for producers and consumers at the market equilibrium price. Hence, in the short run electricity is a commodity, freely flowing into the transmission grid from selected generators and out of the grid to the willing customers. Every half hour, customers pay and generators receive the short-run marginal-cost (SRMC) price for the total quantity of energy supplied in that half hour. Everyone pays or receives the true opportunity cost in the short run. Payments follow in a simple settlement process.

The control of dispatch at the margin does not imply that market participants must relinquish control of their generating plants. Quite the contrary. In a coordinated spot market with discretionary bids, the market participants have the choice of bidding so as to effectively guarantee that their plant runs. This implies that it is an easy matter to accommodate self-scheduling for those who prefer this option. However, the opportunity cost of self-scheduling is a loss of the profits from buying back power when it is cheaper than self-generation, or selling excess output when the price is high. In the end, therefore, there would be natural incentives to bid closer to marginal cost. But it would not be required.

Transmission Congestion

This overview of the short-run market model is by now familiar and found in operation in many countries. However, this introductory overview conceals a critical detail that would be relevant for transmission pricing. Not all power is generated and consumed at the same location. In reality, generating plants and customers are connected through a largely free-flowing grid of transmission and distribution lines.

In the short-run, transmission too is relatively simple. The grid has been built and everyone is connected with no more than certain engineering requirements to meet minimum technical standards. In this short-run world, transmission reduces to nothing more than putting power into one part of the grid and taking it out at another. Power flow is determined by physical laws, but a focus on the flows, whether on a fictional contract path or on more elaborate allocation methods, is a distraction. The simpler model of input somewhere and output somewhere else captures the necessary reality. In this simple model, transmission complicates the short-run market through the introduction of losses and possible congestion costs.

Transmission of power over wires encounters resistance, and resistance creates losses.

Hence the marginal cost of delivering power to different locations differs at least by the marginal effect on losses in the system. Incorporating these losses does not require a major change in the theory or practice of competitive market implementation. Economic dispatch would take account of losses, and the market equilibrium price could be adjusted accordingly. Technically this would yield different marginal costs and different prices, depending on location, but the basic market model and its operation in the short-run would be preserved.

Transmission congestion has a related effect. Limitations in the transmission grid in the short run may constrain long-distance movement of power and thereby impose a higher marginal cost in certain locations. Power will flow over the transmission line from the low cost to the high cost location. If this line has a limit, then in periods of high demand not all the power that could be generated in the low cost region could be used, and some of the cheap plants would be "constrained off." In this case, the demand would be met by higher cost plants that absent the constraint would not run, but due to transmission congestion would be then "constrained on." The marginal cost in the two locations differs because of transmission congestion. The marginal cost of power at the low cost location is no greater than the cost of the cheapest constrained-off plant; otherwise the plant would run. Similarly, the marginal cost at the high cost location is no less than the cost of the most expensive constrained-on plant; otherwise the plant would not be in use. The difference between these two costs, net of marginal losses, is the congestion rental.

This congested-induced marginal-cost difference can be as large as the cost of the generation in the unconstrained case. If a cheap coal plant is constrained off and an oil plant, which costs more than twice as much to run, is constrained on, the difference in marginal costs by region is greater than the cost of energy at the coal plant. This result does not depend in any way on the use of a simple case with a single line and two locations. In a real network the interactions are more complicated, with loop flow and multiple contingencies confronting thermal limits on lines or voltage limits on buses, but the result is the same. It is easy to construct examples where congestion in the transmission grid leads to marginal costs that differ by more than 100% across different locations.

If there is transmission congestion, therefore, the short-run market model and determination of marginal costs must include the effects of the constraints. This extension presents no difficulty in principle. The only impact is that the market now includes a set of prices, one for each location. Economic dispatch would still be the least-cost equilibrium subject to the security constraints. Generators would still bid as before, with the bid understood to be the minimum acceptable price at their location. Customers would bid also, with dispatchable demand and the bid setting the maximum price that would be paid at the customer's location. The security-constrained economic dispatch process would produce the corresponding prices at each location, incorporating the combined effect of generation, losses and congestion. In terms of their own supply and demand, everyone would see a single price, which is the SRMC price of power at their location. If a transmission price is necessary, the natural definition of transmission is supplying power at one location and using it at another. The corresponding transmission price would be the difference between the prices at the two locations.

This same framework lends itself easily to accounting extensions to explicitly include bilateral transactions. Here market participants prefer to schedule point-to-point transmission rather than explicitly buy and sell through the spot market. The bilateral schedules would be provided to the system operator. Those not scheduled would bid into the pool-based spot market. This is often described as the "residual pool" or "net pool" approach. For market participants who wish to schedule transmission between two locations, the opportunity cost of the transmission is just this transmission price of the difference between spot prices at the two locations. This short-run transmission usage pricing, therefore, is efficient and non-discriminatory. In addition, the same principles could apply in a multi-settlement framework, with day-ahead scheduling and real-time dispatch. These extensions could be important in practice, but would not fundamentally change the outline of the structure of electricity markets.

This short-run competitive market with bidding and centralized dispatch is consistent with economic dispatch. The locational prices define the true and full opportunity cost in the short run. Each generator and each customer sees a single price for the half hour, and the prices vary over half hours to reflect changing supply and demand conditions. All the complexities of the power supply grid and network interactions are subsumed under the economic dispatch and calculation of the locational SRMC prices. These are the only prices needed, and payments for short-term energy are the only payments operating in the short run, with administrative overhead covered by rents on losses or, if necessary, a negligible markup applied to all power. The system operator coordinates the dispatch and provides the information for settlement payments, with regulatory oversight to guarantee comparable service through open access to the pool run by the system operator through a bid-based economic dispatch.

With efficient pricing, users have the incentive to respond to the requirements of reliable operation. Absent such price incentives, choice would need to be curtailed and the market limited, in order to give the system operator enough control to counteract the perverse incentives that would be created by prices that did not reflect the marginal costs of dispatch. A competitive market with choice and customer flexibility depends on getting the usage pricing right.

Long-Run Market Contracts

With changing supply and demand conditions, generators and customers will see fluctuations in short-run prices. 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, 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 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 operation of the 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. In a competitive market, 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, 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. In this circumstance, except for the small effect of losses, 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 five cents. On the half hour, if the spot price is six cents, the customer buys power from the coordinated spot market at six cents and the generators sells power for six cents. Under the contract, the generator owes the customer one cent for each of the 100 MW over the half hour. In the reverse case, with the spot price at three cents, the customer pays three cents to the system operator, which in turn pays three cents to the generator, but now the customer owes the generator two cents for each of the 100 MW over the half hour.

In effect, the generator and the customer have a long-term contract for 100 MW at five cents. The contract requires no direct interaction with the system operator other than for the continuing short-run market transactions. But through the interaction with system operator, 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 secondary market, namely the coordinated spot market, where extra power is purchased or sold at the spot price. Similarly for the generator, there is an automatic 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 arithmetic guarantees that the average price is still five cents. Furthermore, with the contract fixed at 100 MW, rather than the amount actually produced or consumed, the long-run average price is guaranteed without disturbing any of the short-run incentives at the margin. Hence the long-run contract is compatible with the short-run market.

The price of the generation contract would depend on the agreed reference price and other terms and conditions. Generators and customers might agree on dead zones, different up-side and down-side price commitments, or anything else that could be negotiated in a free market to reflect the circumstances and risk preferences of the parties. Whether generators pay customers, or the reverse, depends on the terms. However, the system operator need take no notice of the contracts, and have no knowledge of the terms.

In the presence of transmission congestion, 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 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 would be needed.

Transmission congestion in the short-run market raises another related and significant matter for the system operator. 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. If the system operator retains the benefits from congestion rentals, this incentive would work contrary to the goal of an efficient, competitive electricity market.

A convenient solution to both problems--providing a price hedge against locational congestion differentials and removing the adverse incentive for system operator--is to re-distribute the congestion revenue through a system of long-run transmission congestion contracts operating in parallel with the long-run generation contracts. 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 a transmission congestion contract that provides compensation for differences in prices, in this case for differences in the congestion costs between different locations across the network.

The transmission congestion contract for compensation would exist for a particular quantity between two locations. The generator in the example above might obtain a transmission congestion contract 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 congestion rental. Hence, if a transmission constraint caused prices to rise to six cents at the customer's location, but remain at five cents at the generator's location, the one cent difference would be the congestion rental. The customer would pay the system operator six cents for the power. The system operator would in turn pay the generator five cents for the power supplied in the short-run market. As the holder of the transmission congestion contract, the generator would receive one cent for each of the 100 MW covered under the transmission congestion contract. This revenue would allow the generator to pay the difference under the generation contract so that the net cost to the customer is five cents as agreed in the bilateral power contract. Without the transmission congestion contract, the generator would have no revenue to compensate the customer for the difference in the prices at their two locations. The transmission congestion contract completes the package.[79]

When only one generator and one customer are involved, this sequence of exchanges

---

[79] The description here ignores the effect of losses, but this simplification is not important to the argument.

under the two types of contracts may seem unnecessary. However, in a real network with many participants, the process is far less obvious. There will be many possible transmission combinations between different locations. There is no single definition of transmission grid capacity, and it is only meaningful to ask if the configuration of aggregated transmission flows is feasible. However, the net result would be the same. Short-run incentives at the margin would follow the incentives of short-run opportunity costs, and long-run contracts would operate to provide price hedges against specific quantities. The system operator coordinates the short-run market to provide economic dispatch. The system operator collects and pays according to the short-run marginal price at each location, and the system operator distributes the congestion rentals to the holders of transmission congestion contracts. Generators and customers make separate bilateral arrangements for generation contracts. Unlike with the generation contracts, the system operator's participation in coordinating administration of the transmission congestion contracts is necessary because of the network interactions, which make it impossible to link specific customers paying congestion costs with specific customer receiving congestion compensation. If a simple feasibility test is imposed on the transmission congestion contracts awarded to customers, the aggregate congestion payments received by the system operator will fund the congestion payment obligations under the transmission congestion contracts.[80] Still, the congestion 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 transmission congestion contracts define payment obligations that guarantee protection from changes in the congestion rentals.

The transmission congestion contract can be recognized as equivalent to an advantageous form of transaction with controllable lines and the associated point-to-point “physical” transmission rights. With controllable lines the system would be simpler, and the flows across the line would come closer to the expansion of capacity. However, even with controllable line the full increase in capacity in the larger system might be more than the power that flowed across the line. The ability to respond in emergencies or provide voltage support might relieve other constraints and provide additional transmission congestion contracts beyond the strict flows across the line. The transmission congestion contract provides a more general solution, with or without the benefits of a controllable line.

Were it possible to define usage of the transmission system in terms of physical rights, it would be desirable that these rights have two features. First, they could not be withheld from the market to prevent others from using the existing transmission grid. Second, they would be perfectly tradable in a secondary market that would support full reconfiguration of the patterns of network use at no transaction cost. This is impossible with any known system of physical transmission rights that parcel up the transmission grid. However, in a competitive electricity market with a bid-based, security-constrained economic dispatch, transmission congestion contracts are equivalent to just such perfectly tradable transmission rights. Hence we can describe transmission congestion contracts either as financial contracts for congestion rents or as perfectly tradable physical

[80] The feasibility test is simple when the transmission congestion contracts are defined as forward contract obligations. If the transmission congestion contracts are defined as options, the feasibility test is well defined but can be more complicated. This difference between obligations and options applies equally to physical and financial transmission rights. For a further discussion, see 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).

transmission rights.

If the transmission congestion contracts have been fully allocated, then the system operator will be simply a conduit for the distribution of the congestion rentals. The operator would no longer have an incentive to increase congestion rentals: any increase in congestion payments would flow only to the holders of the transmission congestion contracts. The problem of supervising the dispatch monopoly would be greatly reduced. And through a combination of generation contracts and transmission congestion contracts, participants in the electricity market could arrange price hedges that would provide the economic equivalent of a long-term contract for specific power delivered to a specific customer.

Further to the application of these ideas, locational marginal cost pricing lends itself to a natural decomposition. For example, even with loops in a network, market information could be transformed easily into a hub-and-spoke framework with locational price differences on a spoke defining the cost of moving to and from the local hub, and then between hubs. This would simplify without distorting the locational prices. A contract network could develop that would be different from the real network without affecting the meaning or interpretation of the locational prices.

With the market hubs, the participants would see the simplification of having a few hubs that capture most of the price differences of long-distance transmission. Contracts could develop relative to the hubs. The rest of the sometimes important difference in locational prices would appear in the cost of moving power to and from the local hub. Commercial connections in the network could follow a configuration convenient for contracting and trading. The separation of physical and financial flows would allow this flexibility.

The creation or elimination of hubs would require no intervention by regulators or the system operator. New hubs could arise as the market requires, or disappear when not important. A hub is simply a special node or portfolio of nodes within a zone. The system operator still would work with the locational prices, but the market would decide on the degree of simplification needed. However, everyone would still be responsible for the opportunity cost of moving power to and from the local hub. There would be locational prices and this would avoid the substantial incentive problems of averaging prices.

Long-Term Market Investment

Within the contract environment of the competitive electricity market, new investment occurs principally in generating plants, customer facilities and transmission expansions. In each case, corresponding contract-right opportunities appear that can be used to hedge the price uncertainty inherent in the operation of the competitive short-run market.

In the case of investment in new generating plants or consuming facilities, the process is straightforward. Under the competitive assumption, no single generator or customer is a large part of the market, there are no significant economies of scale, and there are no barriers to entry of efficient plants. Generators or customers can connect to the transmission grid at any point subject only to technical requirements defining the physical standards for hookup. If they choose, new customers or new generators have the option of relying solely on the short-run market, buying and selling power at the locational price determined as part of the half-hourly dispatch. The system operator makes no guarantees as to the price at the location. The system operator only guarantees open access to the coordinated spot market at a price consistent with market equilibrium. The investor takes all the business risk of generating or consuming power at an acceptable price.

If the generator or customer wants price certainty, then new generation contracts can be struck between a willing buyer and a willing seller. The complexity and reach of these contracts would be limited only by the needs of the market. Typically we expect a new generator to look for a customer who wants a price hedge, and for the generators to defer investing in new plant until sufficient long-term contracts with customers can be arranged to cover a sufficient portion of the required investment. The generation contracts could be with one or more customers and might involve a mix of fixed charges coupled with the obligations to compensate for price differences relative to the spot-market price. But the customer and generator would ultimately buy and sell power at their location at the half-hourly price.

If either party expects significant transmission congestion, then a transmission congestion contract would be indicated in any case that would otherwise benefit from point-to-point physical rights were they available. If transmission congestion contracts are for sale between the two points, then a contract can be obtained from the holder(s) of existing rights. Or new investment can create new capacity that would support additional transmission congestion contracts. The system operator would participate in the process only to verify that the newly created transmission congestion contracts would be feasible and consistent with the obligation to preserve any existing set of transmission congestion contracts on the existing grid. Unlike the ambiguity in the traditional definition of transmission transfer capacity, there is a direct test to determine the feasibility of any new set of transmission congestion contracts for compensation--while protecting the existing rights--and the test is independent of the actual loads that may develop. Hence, incremental investments in the grid would be possible anywhere without requiring that everyone connected to the grid participate in the negotiations or agree to the allocation of the new transmission congestion contracts.

This happy resolution of the puzzle of transmission expansion and pricing through voluntary market forces alone is subject to at least two other important caveats. First, there still may be market failures even with the definition of a workable set of equivalent property rights. For example, with many small market participants, each benefiting a little from a large transmission investment, the temptation to free-ride on the economies of scale and scope may create a kind of prisoner's dilemma. Everyone would be better off sharing in the investment, but the temptation to free ride and avoid paying for the expense may overcome any ability to form a consortium or negotiate a contract. It may be that the investment could not go forward in a timely manner, at the right scale, or at all, without some regulatory entity that can mandate payment of the costs.[81] In this case, however, the task should be simplified by the ability to simultaneously allocate the benefits in the form of a share of the transmission congestion contracts. The market could take care of many, perhaps most, investments, and the regulatory option would be easier to implement when needed.[82]

A related problem could appear in the circumstances where the pattern of transmission

---

[81] This situation appears to be what is described often as investments for reliability. However, with price-responsive demand and security-constrained economic dispatch, there is in principle no difference in reliability investments and economic investments. The only difference created by the investment would be in the economic benefits of the actual dispatch.

[82] William W. Hogan, "Market-Based Transmission Investments and Competitive Electricity Markets," Center for Business and Government, Harvard University, August 1999.

use was so uncertain and the network so interconnected that no set of point-to-point rights would be capable of capturing enough of the economic benefits of grid investment. This would be true, of course, for both physical rights were they possible, as well as for the transmission congestion contracts. In effect, there would be significant economies of scope in transmission investment that would go well beyond the benefits of any reasonable patterns of point-to-point rights. If the benefits could not be assigned, then the market-based investments would not follow.

Second, operation of voluntary market forces would have little sway in the allocation of the costs for an existing transmission grid that already provides open access. The costs are sunk, and typically the sunk costs of the wires exceed the transmission congestion opportunity costs of using the grid. This is due, in large part, to the effects of the economies of scale. Hence, given the choice of paying the sunk costs but avoiding the congestion costs, versus avoiding the sunk costs while using the system and paying the continuing cost of congestion, most users would prefer the latter. If the sunk costs are to be recovered in prospective payments, therefore, there must be some form of requirement to pay these costs as a condition for using the grid. The resulting access charges would be the functional equivalent of the contract payments for new investment.

This same basic system has been in operation for more than a year under the PJM Interconnection, where the financial transmission rights are labeled as fixed transmission rights (FTR).[83] In the New York ISO structure that began operations in November 1999, the same basic model applies with equivalent transmission rights called transmission congestion contracts (TCC).[84] In the proposed New England ISO structure, the term of art is a financial congestion right (FCR).[85] There are slight differences in all these approaches, but they stand on the bedrock of a coordinated spot market, implement through a bid-based, security-constrained economic dispatch with locational prices. This supports a high degree of choice by market participants, and is the only known model that provides these benefits in a framework to support competitive electricity markets.

Getting the Prices Right

The importance of the bid-based, security-constrained economic dispatch model is illustrated succinctly by the experience in PJM.[86] The short version of a longer story is that PJM conducted something like a natural experiment with alternative approaches to transmission access and congestion management. The full implementation of the current model in PJM, as outlined above, was embraced starting in April 1998, only after an experiment during 1997 with an

---

[83] The PJM system began operation with FTRs and full locational pricing as of April 1, 1998. Details can be obtained from the web site at www.pjm.com.

[84] Federal Energy Regulatory Commission, New York ISO Ruling, Docket Nos. ER97-1523-000, OA97-470-000 and ER97-4234-000, January 27, 1999. See also, "Order Denying in Part and Granting in Part Rehearing and Clarification and Conditionally Accepting Compliance Filing," Dockets Nos. ER97-1523-003 and -004, OA97-470-004 and -005, and ER97-4234-002 and -003, Washington DC, July 29, 1999.

[85] Federal Energy Regulatory Commission, "Order Accepting Preliminary Congestion Management and Multi-Settlement Systems and Providing Guidance," New England Power Pool, Docket No. ER99-2335-000, Washington DC, July 30, 1999.

[86] For further details, 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.

alternative market model and pricing approach that proved to be fundamentally inconsistent with a competitive market and user flexibility.[87] The experiment made the point in a dramatic way. The important issue is not the total cost of transmission congestion, which may be small on average if the system is used efficiently, and when the cost is often mistakenly dismissed as irrelevant. Rather, the point is the incentives at the margin when the system is constrained. In designing the rules for transmission access and pricing for a competitive market, it matters little how the rules perform when the system is unconstrained. The important question is how the rules deal with the market and participant choices when the system is constrained.

The earlier flawed pricing system allowed market participants the flexibility to choose between bilateral transactions and spot purchases, but treated PJM as a single zone and did not simultaneously present market participants with the costs of their choices. The circumstances created a false and artificial impression that savings of $10 per MWh or more could be achieved simply by converting a spot transaction into a bilateral schedule. Faced with this perverse pricing incentive, market participants responded naturally by scheduling more bilateral transactions than the transmission system could accommodate. In effect, using the wrong prices induced behavior which greatly increased the cost of congestion. Inevitably, in June 1997 the ISO had to intervene by restricting the market and constraining choice to preserve reliability. 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, the ISO had no alternative but to restrict the market.

Even if the total cost of congestion might be modest over a year, a gap of $10 per MWh between the true costs of transmission usage and what participants pay is more than sufficient to get the attention of market participants at the time when it matters most, when the system is constrained. Given the margins in this business, market participants will change their behavior for $1. And the changes in behavior can substantially affect system operations; in fact, the whole point of electricity restructuring is that changes in behavior can affect system operations and lead to different patterns of electricity use and investment.

By contrast, the locational pricing system avoids this perverse incentive. By construction, the locational prices equal system marginal costs. 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.

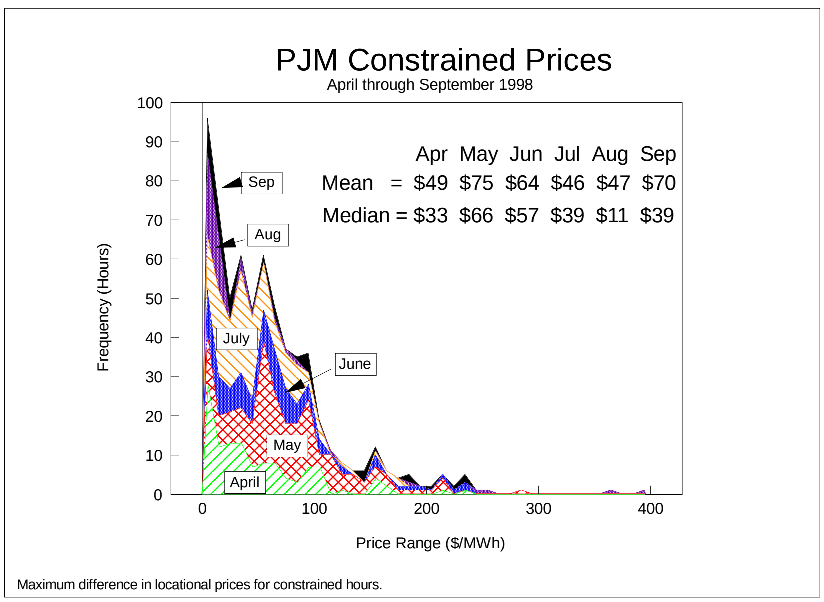
Starting in April 1998--under the bid-based, security-constrained economic dispatch model with locational pricing--the magnitude of transmission congestion became evident for all to see. The system experienced transmission constraints, locational prices separated, and the opportunity cost of transmission was quite large. The lowest locational prices were sometimes negative, reflecting the value of counterflow in the system where it would be cheaper to pay participants to take power at some locations and so relieve transmission constraints. The highest locational prices were larger than the marginal cost of the most expensive plant running, reflecting the need to simultaneously increase output from expensive plants and decrease output from cheap

---

[87] Here the issue is pricing for transmission congestion. The recovery of embedded costs of transmission investment through access charges is a separate matter that is amenable to a zonal approach. Locational pricing has long been available in other markets, such as New Zealand. The PJM case is of interest because of its size, the sharp contrast of the debates, the experiment of trying two different pricing systems, and the availability of the data.

plants, just to meet an increment of load at a constrained location. Over all hours in April 1998, for example, the low price was -$45 at 1500 hours on April 18 at "JACK PS," and the highest price was $232 at 1100 hours on April 16 at "SADDLEBR," both locations being in the Public Service Gas & Electric territory. Over the first year with the locational pricing system, the maximum difference between the lowest and highest contemporaneous prices was $412, at 1100 hours on November 19, reflecting the difference between $322 at "SADDLEBR" and -$91 at "BELLVIL."[88] The second highest difference was $399, at 2000 hours on August 26, reflecting the difference between $437 at "ESAYRE" and $38 at "NYPP-W." This maximum price separation reached the same level as in the relatively unconstrained month of March 1998 before the locational prices were charged, when users could ignore the cost of congestion.[89]

The contemporaneous difference in locational prices, which is the price of transmission usage, has been large quite often. It does not take much of a difference to change behavior when the reported trading margins may be as low as $1 per MWh. If we take the $1 per MWh standard as an arbitrary threshold to define a constrained period, the range of highest to lowest price across locations exceeded the threshold for 119 hours in April, or approximately 17% of the time. As shown in the accompanying figure, the frequency distribution of the price range in constrained hours is skewed, with a median hourly price range at $33 and a mean of $49 in April. When the system is constrained and the market incentives matter the most, the marginal costs of transmission can be large indeed.



The monthly data for May through September, covering the summer peak, reinforce this initial impression. In general, May saw both higher prices and more transmission congestion. The difference between the highest and lowest locational price in May exceeded the $1 threshold for 183 hours, or approximately 25% of the time. As shown in the accompanying figure, the frequency distribution of this congestion price shifted to higher costs. In May, the median of the hourly price ranges doubled to $66 and the mean increased to $75. June was less constrained, exceeding the $1 threshold for 95 hours or 13% of the time. The June median of the hourly price ranges was

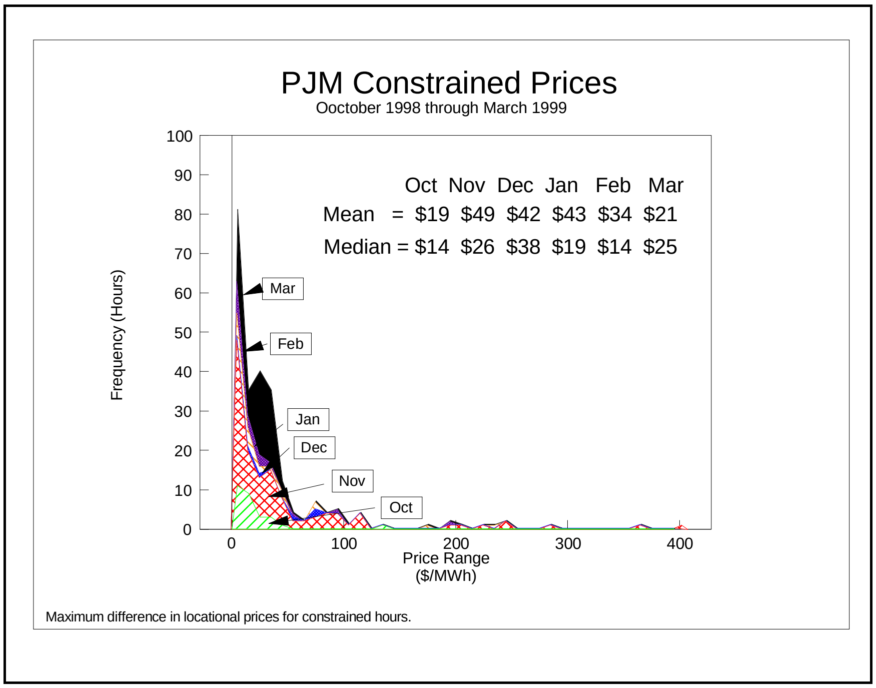
88 In July 1999, the maximum difference in locational prices reached $1700.

89 On March 26, 1998, at 2200 hours, the difference between the highest to the lowest marginal cost was almost $400.

$57 and the mean was $64. During July, constraints appeared more often, as in May, with 151 constrained hours or about 20% of the time. July saw a median hourly price range of $39 and a mean of $46. By contrast, August showed locational constraints only for 48 hours or 7% of the time. The median hourly price range in August was $11 and the mean was $47, reflecting a few hours when the difference between the lowest to the highest price reached almost $400. September was like August, with 46 constrained hours or 6% of the time. However, the average price of congestion was high in September, with a mean of $70 and a median of $39.

The experience of higher unconstrained prices and fewer constrained hours in June, August and September reminds us that the period of peak system load is not necessarily the time of greatest transmission congestion. Transmission congestion reflects an imbalance in the location of load and generation. At peak load, more generation comes on line and may relieve system congestion. In addition, the particular flow of power into the Midwest, reversing the usual direction, tended to unload the transmission constraints during the summer of 1998.

This record of continuing constraints was reinforced by the events in the following months from October 1998 to March 1999. After the heavier loading of the PJM summer, the winter months would be less constrained but the constraints did not disappear. As shown in the accompanying figure, the frequency diagram of price ranges showed that some significant constraints applied. November alone accounted for 105 of the 242 constrained hours over the period. The median price range for the constrained hours in November was $26 and the mean was $49. The corresponding median and mean price ranges of the other months for the hours that the system was constrained appear in the figure.[90]



The evidence shows many things. For example, calculating and reporting the locational prices for each point on the grid are not especially complex tasks, at least for the system operator who has the necessary information available.[91] The prices can be available every five minutes on the Internet. Faced with these prices, the market participants adjust their behavior, just as intended.

---

[90] The numbers of constrained hours for October 1998 through March 1999 were 28, 105, 6, 20, 18, and 65, respectively.

[91] The PJM ISO keeps track of prices at approximately 2000 locations, of which several hundred may have different prices at any moment.

The transition was not painless, especially for those who ignored many warnings and entered into "seller's choice" contracts that gave the seller the maximum theoretical financial advantage for relieving congestion. Presumably, this form of contract will disappear, or be properly priced in the future, and market participants will become more attuned to the use of fixed transmission rights to hedge much of the cost of congestion. But market participants who rely on the spot market, and are not prepared to pay for congestion hedges that fix the cost of transmission in advance, will see price signals that align their incentives with the reality of system operations.

Full locational pricing is fully compatible with a trading system built on a hub-and-spoke framework. The hub becomes a common trading point, and the cost of moving to and from the hub, along the spokes, is just the difference in the locational prices. If the nodal prices are available from the ISO, market participants can define their own hubs. In the PJM case, however, market participants asked the ISO to handle the accounting to create several hubs, of which the western hub has so far developed as the preferred trading point.

The full market response to all these changes is not known because the data are not all in the public domain. However, one information source is a sampling of trader activity reported in the Wall Street Journal.[92] According to these data, the immediate response of the market was to reduce reported spot trading in April of 1998. However, by mid-May of 1998 reported transactions had returned to volumes comparable to those seen just before the new locational pricing system went into effect. Subsequently, and reversing its earlier objection that the nodal pricing market would not be sufficiently liquid, in March of 1999 the New York Mercantile Exchange launched a new futures contract to capitalize on the highly liquid trading market that had developed at the PJM western hub. Apparently inadequate liquidity was not a problem. Even further, the spot and forward markets at the western hub were reported to be so liquid that the futures contract might not be able to compete.[93] Although market liquidity is often vaguely defined and seen only in the eye of the beholder, as reported by these sources the market appears to have adjusted to the new environment within a framework that supports transactions with consistent prices.

The operational problems experienced by the ISO in the year before full locational pricing, where profit driven market participants undermined reliability, did not appear in the year after adoption of full locational pricing. Locational pricing presents profit driven market participants with the right incentives consistent with the true opportunity costs. This same pricing system was applied by PJM for managing inter-regional transmission loading relief. With full locational pricing, the prices reinforce reliability. In addition, the anecdotal evidence suggests that investments in new generation and transmission were being considered with careful attention to the effects of system congestion, just as intended.

In summary, the first year in PJM demonstrated what the system operators knew but which the rest of us need to learn. Our intuition about the magnitude and pattern of transmission congestion is bad, very bad. Despite the earlier pronouncements that transmission congestion would not be significant in PJM, the data show just the opposite. The costs of transmission

---

[92] For example, see "DJ Electricity Price Indexes," Wall Street Journal, June 3, 1998, p. C19.

[93] "The New York Mercantile Exchange will launch an electricity futures contract March 19 at the PJM western hub, one of the most liquid markets in the Eastern grid. ... The PJM hub already features an active and growing over-the-counter forwards market. A liquid hub can have a downside [for the futures contract] given that players are content trading in the OTC, said one Northeast broker." Power Markets Week, February 8, 1999, p. 14.

congestion can be very high, and failure to internalize these costs can completely disrupt the energy market. This is not a mere technical detail. From the perspective of designing market institutions, it is the most important phenomenon. The best way to deal with the reality is as was found in PJM. The system operator must offer a balancing service. Done properly, this amounts to coordinating the spot market. The natural framework is the bid-based, security-constrained economic dispatch with locational prices. The locational prices can be surprising and the differences can be large. It is these large differences that make it especially important to use the prices in charging for transmission congestion. This is the only workable system that can support a non-discriminatory competitive market that allows for participant choice and flexibility. And as experience in PJM and elsewhere shows, it works.

TRANSMISSION ACCESS AND PRICING

The implications for transmission access and pricing affect many aspects of the market rules. [94] The basic outline of the bid-based, security-constrained economic dispatch with locational prices and transmission congestion contracts provides a foundation. It is a foundation that is firmly planted in the Commission's RTO Order and the earlier CRT proposal. If we could move beyond the distracting debates, and avoid the avoidable design flaws for the system operator, we could turn our attentions to an unfinished agenda. The constraints of space permit only an outline of the analysis of these other important matters. Some of the major, interconnected topics include:

- **Transmission Pricing for Fixed Charge Recovery**. Transmission displays large economies of scale, which means that the efficient way to recover embedded costs is in a system that separates fixed and variable charges. Recovery of fixed costs would further distinguish between those designed to recover sunk costs and those that would apply to new investments. There is no reason that sunk cost recovery should be uniform across the grid. This is a point of common confusion. Recognizing that access charges could and should be different at different locations is not the same thing as rejecting a postage stamp rate. For the sunk costs, the license plate approach with different access charges for different regions--now in place, for example, in PJM, California and Australia--would be a preferred alternative, extending into the future. This idea is embodied in the RTO Order.[95] For new investment, fixed charges could be collected under a separate contract with those obtaining rights in the expanded system. This means that locationally different access charges should be the norm, not a temporary transition arrangement. But once the participant pays the access charges, the access should work like a postage stamp covering the entire grid with the right to participate in the coordinated spot market. Importantly, recovery of sunk costs has all the familiar characteristics of the stranded asset problem. In a world of truly open access, market participants would not willingly pay for the sunk costs. Hence, it is important (mandatory) that payments for embedded costs be mandatory.
- **Transmission Rights for Energy**. The locational pricing system for managing transmission usage and congestion is central to the competitive market design. The structure of this pricing

[94] In light of the RTO Order, this revises the similar summary from William W. Hogan, "Restructuring the Electricity market: Institutions for Network Systems," Harvard-Japan Project on Energy and the Environment, Center for Business and Government, Harvard University, April 1999, pp. 16-19.

[95] RTO Order, p. 524.

system supports the definition of transmission rights, incentives for investment, and the ability to rely on competitive market forces. Under a market structure with bid-based economic dispatch and locational pricing, the natural definition of transmission rights is in the form of financial contracts for the collection of congestion rentals. This approach is also recognized in the RTO Order.[96] The rights are FTRs in PJM or TCCs in New York. Furthermore, the rights could also serve as an important element of the incentives and obligations for the grid company. The basic design could be extended to include the explicit treatment of marginal losses and loss payments, as under consideration in New Zealand. Allocation of FTRs through the acquisition of network service creates certain perverse incentives in the re-designation process. An alternative is the development of auction mechanisms for allocating and reconfiguring transmission rights. Such auctions have operated in PJM and been launched in New York.

- **Transmission Requirements for Connection**. There is a distinction between the requirements for new generators and loads to connect to the grid and the requirements for expansion of the grid. The connection rules will be important as both an obligation for the grid company and as an important source of future revenue. At the same time, we should ensure that the connection policies do not become a barrier to entry. In effect, connection for a new generator should require nothing more than would be dictated by the electrical impacts of the plant if it were running and producing no energy, which is always an option.

- **Transmission Expansion Protocols**. Separate from the connection to the grid are the incentives and rules for transmission system expansion in order to increase the capacity for energy trading and to create new FTRs.[97] Here we would distinguish between market driven expansions and expansions dictated by market failure.[98] For market driven expansions, the incentive for investment comes from the market participants who do not wish to pay future congestion costs and seek new FTRs. Or there might develop a mechanism that limited the working capacity of new investments for enough time to justify the return based only on ex post congestion rents. The market participants could approach the existing grid company and arrange for the investment, contract for future fixed payments, and receive the resulting FTRs. Presumably, any existing or new grid company could compete for this business, which could be contestable. For investments where market failure prevents the development of a transmission expansion, there would be a different but parallel decision mechanism. In this case, the grid company might take the initiative in identifying cost-effective expansion options that cannot be undertaken because of the free-rider problem blocking the formation of a sufficiently large coalition of beneficiaries. If the case can be established, the grid company could propose an allocation of costs and benefits. The costs would be collected in a manner similar to sunk costs for the existing grid. The benefits would be distributed in the form of incremental FTRs.

- **Transmission Rights for Capacity Credit**. As long as there is an installed generation capacity requirement and market--as in PJM, New York and New England--there will be a need for a separate system of transmission capacity rights that would be similar in structure to the old notion of physical point-to-point rights. Although these rights would not be connected to the

---

[96] RTO Order, pp. 382-383.

[97] RTO Order, pp. 488-490.

[98] William W. Hogan, "Market-Based Transmission Investments and Competitive Electricity Markets," Center for Business and Government, Harvard University, August 1999, available on the author's web page.

congestion payments, they would be defined according to a simultaneous feasibility condition for deliverability. The rights could be auctioned, tradable, and so on.

- **Transmission and Market Power**. There is a potential interaction between transmission rights and the exercise of generation market power.[99] The basic point is that generators with market power could affect both the profitability of their generation and the value of any transmission rights they may hold.[100]

- **Coordination Across Regions and Transmission Loading Relief**. In a large interconnected grid, the issue of coordination across regions has important implications for the design and use of transmission rights, transmission expansion, and all the other aspects of the transmission business. There is a continuing debate over the procedures for transmission loading relief. A poorly designed TLR mechanism could undermine the market structure and severely reduce the value of the FTRs that can be viewed as the service provided by the wires company. This is sometimes referred to as the "seams" issue, coordinating the operations across regional ISOs in an interconnected grid.[101]

- **Ancillary Services**. Market implementation problems have raised the level of concern about the treatment of the many activities that fall under the heading of ancillary services. Definitions vary, but many of these services could come directly from investments in generation or in the wires business, e.g. with capacitors to provide reactive power support. The policy focus on the energy and capacity markets should be balanced by some further investigation into the developments in ancillary services.[102]

- **Obligations of the Grid Company**. The obligations to be imposed on grid companies have not been fully addressed in the context of the new competitive market designs. The England and Wales case is an exception, with its own problems. Elsewhere, the requirements for the Gridco have not been the focus of policy development. This is an area that could impose potentially large obligations on the Gridco, and where the complexity of the problems provides an opportunity to shift costs. For example, in the absence of well-defined measures of grid capacity, a Gridco might be expected to maintain very large capacity to minimize the possibility that any possible pattern of use could be achieved at low cost. A natural, market-oriented alternative would be to define the product provided by the Gridco in terms of the FTRs and other transmission rights created by the grid, and then to define a set of financial obligations that would be connected, for example, to increased congestion costs. For example, to the extent that the Gridco did not maintain the capacity associated with the FTRs, it could be responsible for any excess congestion cost payments.

- **Incentives for the Grid Company**. The favorite subject of the grid owners is the financial

---

[99] Judith B. Cardell, Carrie Cullen Hitt, William W. Hogan, "Market Power and Strategic Interaction in Electricity Networks," Resource and Energy Economics, 19(1997) 109-137.

[100] RTO Order, pp. 256-257.

[101] For an outline of such coordination measures, see 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, available on the author's web page.

[102] Eric Hirst and Brendan Kirby, "The Functions, Metrics, Costs, and Prices for Three Ancillary Services," Prepared for the Edison Electric Institute, October 1998.

incentives they should enjoy for maintaining the grid and expanding in a cost-effective manner. Such incentives would be appropriate if connected to the framework for obligations. For example, if investment in new grid facilities were contestable, and the obligations of the grid company principally were to stand behind the financial commitments in FTRs and other transmission rights, then a form of light-handed regulation would be possible, with cost-based rates for embedded cost recovery for sunk costs but negotiated market-based payments for some or all new investments. This is an area that has substantial potential, but the ideas are not yet well developed in the policy discussion.

This is not an exhaustive list, with its focus on market operations rather than governance and legal organization. But it does suggest the areas where progress is possible. The key to success would be to build on the fundamental economics of competitive electricity markets, and get the prices right.