DESIGNING MARKET INSTITUTIONS FOR ELECTRIC NETWORK SYSTEMS: REFORMING THE REFORMS IN NEW ZEALAND AND THE U.S.

WILLIAM W. HOGAN

March 13, 2001

Paper for The Utility Convention
Auckland, New Zealand

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

CONTENTS

DESIGNING MARKET INSTITUTIONS FOR ELECTRIC NETWORK SYSTEMS: REFORMING THE REFORMS IN NEW ZEALAND AND THE U.S.

William W. Hogan[1]
March 13, 2001

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. 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. Government and governance mechanisms should keep these principles at the forefront in pursuit of the public interest.

INTRODUCTION

The international experience in designing market institutions has been reflected in the many debates and experiments in the United States. In part, the United States has learned from the best thinking in New Zealand, and recent experience can provide guidance for the market reforms in both countries. The details matter, as is illustrated by examples of both success and failure.

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

---

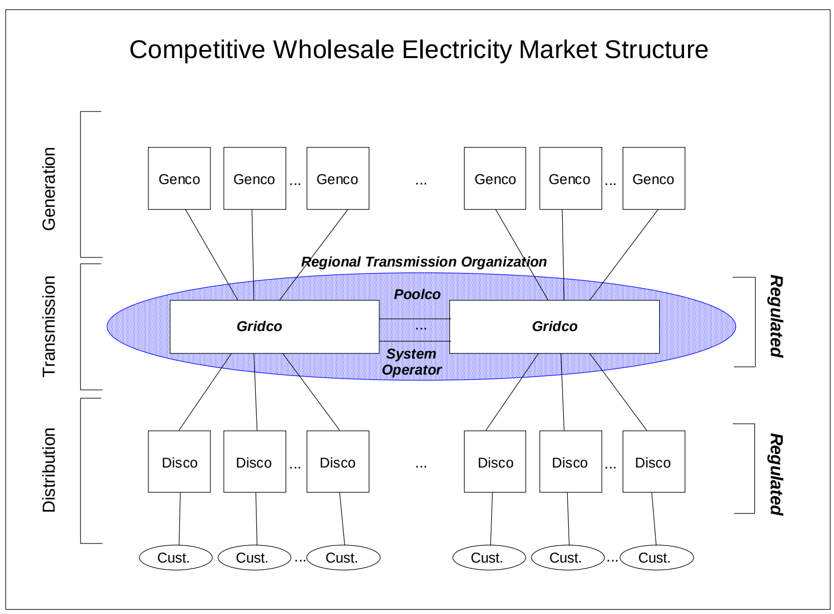
[1] William W. Hogan is the Lucius N. Littauer Professor of Public Policy and Administration, John F. Kennedy School of Government, Harvard University and a Director of the Law and Economics Consulting Group in Navigant Consulting, Inc. This paper draws on work for the Harvard Electricity Policy Group 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., Mirant Corp., National Independent Energy Producers, New England Independent System Operator, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, PJM Office of Interconnection, San Diego Gas & Electric Corporation, Sempra Energy, 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).

be discovered through the magic of the market. 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, have not been easy to fix.

Effective governance mechanisms must address the public interest. An efficient market design can be 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. 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. 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.

THE ESSENTIAL MARKET INGREDIENTS

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. 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 process of restructuring wholesale electricity markets in the United States has added to the extensive worldwide debate about the range of possible and preferred alternatives for organizing regional electricity markets. Most importantly, the Federal Energy Regulatory Commission (FERC) has addressed a wide range of issues in its analysis of and orders for the design of Regional Transmission Organizations (RTO).[2] To signal its importance, the FERC assigned the RTO Order the millennium number. This Millennium RTO Order covers a great deal in fashioning well-designed market institutions to serve the public interest.

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

In the short run, there are critical functions that must be performed by someone. The complex network interactions in an electric grid require that there be an entity that can provide certain critical coordinating services.[3] But the implications that follow from this fact are so contentious that the discussion often becomes confused and the language strained. Here we focus on the activities of this entity as the system operator, no matter what final name we may give it.

The most obvious example of the essential services is in energy balancing. The electric system must maintain continuous aggregate balance of production and consumption. This same balance of inputs and outputs must be coordinated in a way that respects the many limits in the transmission system. Hence, not only must the aggregate inputs and outputs conform to the electrical laws that govern the interconnected grid, but the locational pattern of power production and use must honor these same laws in order to manage the flow of power within the limits of the transmission system.[4] 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 policy should be to require good design for the functions of the system operator.

A central problem appears in designing the design process. Experience indicates that reliance on voluntary agreements among market participants is not likely to be successful. Some

---

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

3 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.

4 RTO Order, pp. 423-424.

problems, like dividing the pie, are largely political and voluntary agreement would be natural. But other problems, like designing bridges, dictate a need for careful consideration of how the pieces fit together and what is in the public interest. Electricity market design is more like the latter than the former.

The 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 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.

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

To do anything else would be to decide on providing the essential coordination services in a way that would be inconsistent with the fundamental goals of electricity restructuring and inconsistent with the basic principle of designing market institutions to support the public interest. As a matter of good public policy, we should not have an interest in market designs that raise costs and decrease the real flexibility of market participants.[5]

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.

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.[6] The same model has been adopted in New York,[7] and embraced as a reform in New England.[8] Likewise, the problems that arise when we do anything else are apparent in various experiments where putative simplifications produced predictable problems.[9]

---

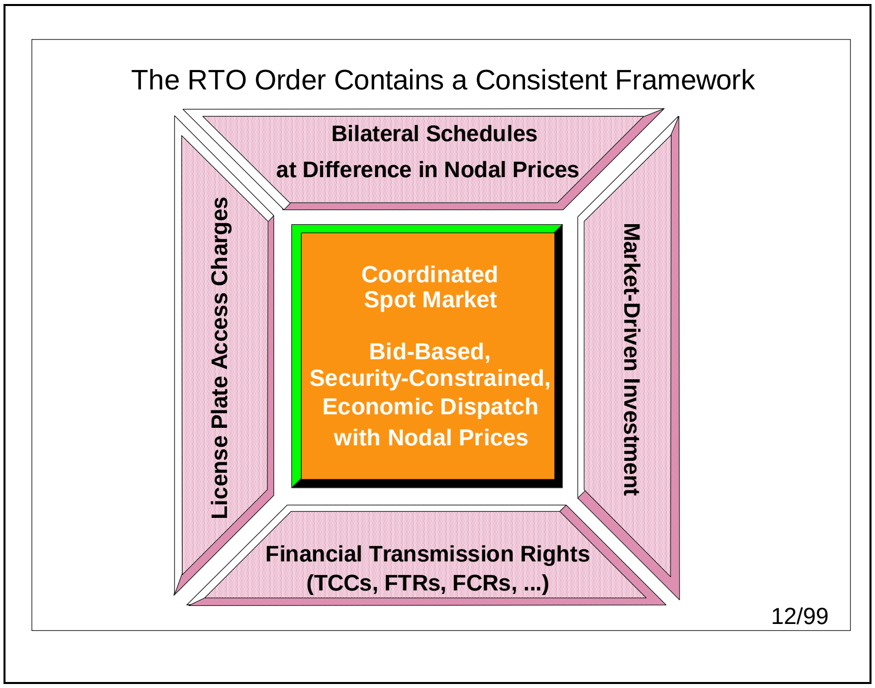
5 Larry Ruff, "Competitive Electricity Markets: Why They Are Working And How To Improve Them," National Economic Research Associates, May 12, 1999.

6 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.

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

8 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.

9 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.



From the perspective of design of institutions, the most important theme running through the Millennium Order's discussion of these characteristics and functions is the prominence of markets as the means for achieving the many goals of electricity restructuring. The key element is in the recognition of the importance of a coordinated spot market. In the Millennium Order this appears principally in the discussion of the balancing market. In particular, "[r]eal-time balancing is usually achieved through the direct control of select generators (and, in some cases, loads) who increase or decrease their output (or consumption in the case of loads) in response to instructions from the system operator."[10] 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."[11]

Further, the FERC requires that everyone be able to participate in this coordinated spot market, at the efficient, and necessarily locational or nodal, prices: "The market mechanisms must accommodate broad participation by all market participants, and must provide all transmission customers with efficient price signals regarding the consequences of their transmission usage decisions."[12] 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."[13]

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, giving rise to the creation of financial transmission rights.[14] Payment for the existing grid would appear in part

---

10 RTO Order, p. 635.

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

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

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

14 RTO Order, pp. 382-383.

as access charges, including the use of the "license plate" approach with region-specific access charges.[15]

These are the most important elements.[16] 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.

CALIFORNIA MELTDOWN

The success of wholesale power markets built on a coordinated spot market stands in contrast to the cascading failure of a major market in the United States that has so far rejected this approach. At the end of 2000, a power crisis in California had laid bare the dangers of ignoring the fundamentals of how power systems operate while creating a caricature of a retail electricity market with a dangerous combination of bad economic theory and worse political economy practice.[17] In the event, wholesale prices surged and stayed above $150 per MWh while retail prices for the same energy were limited to $65. The system soon fell apart, and "deregulation" was pronounced dead.

The bad economic theory was a full embrace of the objective of creating a market for middlemen, no matter what the cost. In California, the pool approach to a coordinated spot market was explicitly rejected in preference to a complicated trading regime. Given the inevitable requirements for coordination, this produced an expanding collection of arcane rules to prevent what was natural by making the coordination process ever harder to use, all in the interest of supporting separate exchanges and marketers. For example, the California system operator was explicitly precluded from providing a least-cost combination of balancing services. Since the operator still had to provide balancing services, these were required to be inefficient and expensive, in order to create more business for the middlemen.

The bad political economy appeared in the process that produced the compromise rules for the California market. Key parts of the decentralized theory would have retail customers face market prices, and the old monopolies would be precluded from anything other than providing distribution services. The political process produced the second rule, and precluded old monopolies from participating in the market. But retail customers were protected from the market price through a fixed price for retail sales. In the event, this eliminated the entry opportunity for the marketers by eliminating the need their principal service (price hedging) and left the old monopolies buying at a variable wholesale market price and selling at a fixed retail price. This proved to be an explosive combination.

---

15 RTO Order, p. 524.

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

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

The compounding failures in the market design accumulated from its inception in 1998 until, at the end of 1999, Federal regulators pronounced the California design as "fundamentally flawed."[18] There then began an intense process to rethink the market design from first principles. The process was made more difficult by the rear-guard action of stakeholder interests that benefited from or created the flawed design.

In the event, the redesign effort was blown aside in the summer of 2000 when the explosive combination of variable wholesale prices and fixed retail prices confronted the spark of a suddenly tight market. Bad luck collided with bad policy. There had been little addition to generating capacity for more than a decade. Low water reservoirs behind power dams combined with higher natural gas prices and tighter environmental conditions. An unexpected surge in demand from economic growth hit the inefficient market and produced unprecedented price increases. Soon the old monopolies were selling power retail for a small fraction of what they paid to acquire the power in the spot wholesale market. Bankruptcy loomed and supply could no longer be assured. Even those who predicted problems were surprised at the scope and speed of the policy disaster.

The Fatal Flaw

California built its market design on a flawed premise. It is a commonplace that electric systems are both complicated and highly interdependent. Over short horizons of a day or less, generating facilities must work through the transmission network to provide the multiple products of energy, reserves and ancillary services. The same generating facilities must provide all of these products, in the right amounts, and with very limited tolerances. The simple physical reality dictates that these services must, in the end, be coordinated by a system operator. There is no other choice available with our current technology, and every electric system has such a system operator.

The flawed premise of the California market design was that this inescapable reality could be ignored or minimized in an effort to honor a faith in the ability of markets to solve the problems of coordination. Worse yet, the design of the California market embraced the notion that what little the system operator would do should be done inefficiently in order to leave even more coordination problems for the market to solve. This was an unprecedented experiment in markets that did not work in theory.[19] We now know that it did not work in practice either.

The failed experiment is at the root of many of the market defects. And the root is deep. The principles have been embodied as part of the so-called "four pillars" of the California market design.[20] Throughout the review of the market design in the intensive process that began when the FERC identified the "fundamentally flawed" congestion management system, the California

---

[18] Federal Energy Regulatory Commission, FERC Docket No. ER00-555-000 January 7, 2000. The resulting CAISO process began with Congestion Management Reform (CMR) and expanded into redefinition of the acronym as Comprehensive Market Redesign. See www.caiso.com for further details.

[19] William W. Hogan, "A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy," Electricity Journal, December 1995.

[20] “Congestion Management Reform,” presentation by the California ISO, March 17, 2000.

Independent System Operator (CAISO) has reflected the will of some stakeholders that above all else the four pillars must be preserved.[21]

These four pillars include:

- The design should "separate the forward energy markets from the ISO forward transmission market."
- The design should "use second-price auction and marginal cost pricing for transmission."
- The design should "utilize the principle of market separation," such as requiring the ISO to preserve balanced schedules for each scheduling coordinator, notwithstanding the ISO's need to adjust these schedules to manage congestion and balance the system.
- The design should "use zonal congestion design where prices within a zone are close enough to use one price for the whole zone."

Only the second principle, to use marginal cost pricing, has a basis in theory or been shown to be workable in practice. Unfortunately, many of the perverse incentives in the California market arise precisely because the ISO is not allowed to apply even this principle consistently. At the same time, the remaining three pillars stand in opposition to the reality of how electric systems must work.

Separating forward energy markets from the ISO's forward transmission markets is a mistake. Over short horizons, there is no distinction between energy dispatch and transmission use. Once we know the dispatch of plants needed to produce energy to meet load, the use of the transmission system is determined. It is a fallacy that these can be determined separately, or that these functions do not have to be carefully integrated to achieve both economic efficiency and reliable operation. Furthermore, this same flawed market separation principle leads to explicit prohibitions of economic dispatch. The separation of day-ahead transmission and energy markets creates problems that could be and have been avoided elsewhere.

Similarly, the principle of market separation that gives rise to the requirement for individually balanced schedules imposes constraints on operations that are designed solely to create opportunities for otherwise unnecessary transactions for the California Power Exchange (PX) and other scheduling coordinators. Aggregate balancing is required by the physics. But individual balancing is not required, often not efficient, and sometimes not even possible. The restriction is entirely artificial and makes it harder for the ISO to coordinate the market. Moreover, the restriction appears likely to increase the capacity shortage in the California market by increasing the CAISO's demand for capacity (to provide regulation) and requiring market participants to withhold capacity from the energy markets in order to provide adjustment bids.

Likewise, the zonal pricing system defines a requirement that should not be a requirement at all given the conditions in its definition. If the (true) prices in a zone were "close enough," there would be no need to convert them to one price. Furthermore, we know by now that the implied simplification of the zonal system was a mirage, and its implementation requires more and more complex contortions to counteract its perverse incentives. The real impact of zonal

---

[21] See, for example, California ISO, Congestion Management Reform Recommendations, Appendix E, July 28, 2000.

aggregation is to convert (true) prices that are not close into a single price that gives the wrong incentives just when incentives matter most.

These ill-advised pillars have trapped California in a box that excludes meaningful market reform. The requirement for individually balanced schedules, rather than a collectively balanced system, serves no good public policy purpose. The prohibition against economic dispatch in real time necessarily reduces efficiency and forecloses a market-based option that is fundamental to workable markets in other systems. The continued pursuit of "simplified" zonal designs, that are truly complicated in practice, reflects the perverse philosophical commitment to preventing the CAISO from doing well what it must do of necessity. The initial complete and still partial separation of markets for energy, reserves and other ancillary services imposes demands on market participants, and on the supply of generating capacity, that could be alleviated easily in the use of a combined optimization that only the CAISO could perform.

There is an understandable focus on high prices and efforts to mitigate the impact on California consumers. Near-term efforts to define just and reasonable prices receive immediate attention, often at the expense of efforts to correct the underlying flaws in the market. But even here the design flaws intrude. They confound diagnosis and treatment of the market ills in California. Initially, high prices in California were seen as *prima facie* evidence of the exercise of market power. However, closer examination of the structure of the market and its rules reveals a more complicated story that implicates the interaction of bad market design and shortage as at least a prominent feature of the California experience.[22] Without the fundamental reforms in market design, it may be impossible to separate the effects of market power from these other elements. And without a better diagnosis, it is hard to know what treatments to prescribe to mitigate market power, or even if market power is a part of the problem. Furthermore, if the real problems have been a combination of a shortage of capacity and high cost energy, market reform may be essential to achieving just and reasonable prices.

The Regulator's Reforms for California

The FERC has reviewed the accumulated experience in California and produced a series of proposed actions for the immediate future and for more fundamental reform. The initial actions include: [23]

- the elimination of the requirement that the three investor-owned utilities (IOUs) -- Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SoCal Edison), and San Diego Gas & Electric Company (SDG&E) -- must sell into and buy from the PX;

---

[22] Scott M. Harvey and William W. Hogan, "Issues in the Analysis of Market Power in California," October 27, 2000. Federal Energy Regulatory Commission, "Staff Report to the Federal Energy Regulatory Commission on Western Markets and the Causes of the Summer 2000 Price Abnormalities," Part 1 of Staff Report on U.S. Bulk Power Markets, November 1, 2000.

[23] Federal Energy Regulatory Commission, "Order Proposing Remedies for California Wholesale Electric Markets," Docket No. EL00-95-000, Washington, DC, November 1, 2000, p. 5.

- the addition of a penalty charge for deviations in scheduling in excess of five percent of an entity's hourly load requirements and the disbursement of penalty revenues to the loads that scheduled accurately;
- the establishment of independent, non-stakeholder Governing Boards for the PX and the ISO;
- the establishment of generation interconnection procedures; and
- a new form of "soft" price cap at $150.

Further, the FERC identified a number of structural reforms that must be addressed, including:[24]

- the submission of a congestion management redesign proposal;
- possible changes to the auction mechanisms;
- improved market monitoring and market mitigation strategies;
- demand response programs by the ISO and Scheduling Coordinators;
- elimination of the requirement for balanced schedules; and
- new approach to reserve requirements.

This is an ambitious agenda, pointing towards undertaking a comprehensive redesign of the entire California market structure.

Governance

The California governance arrangements have failed to meet the basic test of operating success. The governance mechanism that produced the flawed initial market design evolved into the stakeholder boards of the CAISO and the PX. As is now clear, this governance mechanism has been unable to correct, or even acknowledge, its initial mistakes. The FERC has concluded that California needs a new, more independent, governance mechanism. This is an important step that will have major impacts both inside and outside California.

Market Separation

The flaw of market separation received attention from the FERC in its direction regarding the functions of the CAISO and the PX. The CAISO should be given the clear responsibility to run an efficient day-ahead and real-time market, in support of an efficient competitive market. Pricing rules in each market should be based on standard marginal cost principles and be consistent across markets. Any attempt to straddle the four pillars and maintain market separation is bound to fail. There should be an unambiguous decision and direction to give the CAISO the responsibility to operate an integrated system for day-ahead and real-time scheduling, balancing, congestion management, ancillary services, reserves, and so on, recognizing that these and their associated pricing must be parts of an integrated whole.

---

[24] Federal Energy Regulatory Commission, "Order Proposing Remedies for California Wholesale Electric Markets," Docket No. EL00-95-000, Washington, DC, November 1, 2000, p. 5.

Forward Contracting

Freeing utilities from restrictions on forward contracting is a move in the right direction. In a real market, there would be no such restrictions. The arguments for the restrictions in the first place were at best problematic. Whatever the original merits, the arguments depended in part upon other market reforms that would allow for vigorous competition to serve retail loads. These other reforms were not put in place. In addition, the well documented effect of the rate freeze and stranded asset recovery mechanism created the worst possible combination of small customers left *de facto* without access to retail suppliers who could provide price stability, and utilities precluded from providing any hedging services.

Soft Price Cap

The soft price cap proposal is novel and raises many new issues. Essentially the soft price cap appears to be an attempt to straddle two auction price regimes, with market-clearing prices applying below $150 and pay-as-bid systems applying above $150. Below $150 it would seem that any price would be acceptable. Above $150, there would at least be requirements for further review by the FERC and possible refunds.

As with any price cap, the incentives run against the operation of markets and make the mechanism a source of complication in achieving a transition to a more market-like mechanism. It would be especially problematic for prospective new entrants. Consider a competitive existing generator with production costs below but opportunity costs above $150. The opportunity costs should set a floor on its bid in a competitive market. Under a truly "soft" price cap, the risk for such an entity of bidding above $150 would be limited to the cost of filing and review by the FERC, plus the possibility that a refund may be required to return its short-run operating profits in excess of $150. There would be no rational reason not to bid the supplier's opportunity costs, as the worst case outcome would be no worse than if it did not try to capture its opportunity costs in its bid. By contrast, consider the new generator that needs a significant number of hours with revenue above $150 to justify the fixed costs of building a plant and entering the market. Any new generator (or the generator contemplating closing a plant, or a generator contemplating an investment to improve generating performance or reduce NOx emissions) would face a larger maximum risk and would have to evaluate the chance that it would make a cash investment and then not recover its required return. In this case, it is not simply a matter of failing to capture its opportunity costs and being no worse off than if it had not tried, because the ability to capture opportunity costs may have provided the basis for an investment that would be sunk and would fail to recover its cost of capital. It is easy to imagine that this soft price cap would have almost the same effect as a hard price cap for such entrants, namely discouraging new entry. Given the short supply situation, this would be just the wrong incentive.

Auction Mechanisms

The FERC expressed an interest in the possible benefits of switching to a pay-as-bid auction format rather than the originally intended design of a uniform price auction. Electricity markets that rely on uniform price auctions to clear markets exploit a simple argument based on the law of one price. The law of one price says that in a decentralized market for a homogeneous commodity, trade will tend to converge towards a common market-clearing price. In the case of electricity, where decentralized trading is foreclosed in the final day-ahead and real-time markets, this convergence is not possible and the simple approach is to use what the market would produce if only there were enough time and no transaction costs.

Whenever these uniform price electricity markets encounter trouble for any reason, someone notices that market participants are responding to the incentives of the uniform price auction by bidding something below the market-clearing price. They then leap to the *non sequitur* that paying the bid rather than the market-clearing price would somehow reduce average prices. A moment's reflection would suggest that the same market participants who respond to the incentives of the uniform price auction would also respond to the incentives of the pay-as-bid auction. Now the incentive would be to bid the market-clearing price.

The result could be the same price and revenue flows as under the uniform price auction.[25] This assumes, however, that there would be no uncertainty and no transaction costs. In the presence of uncertainty and transaction costs, there will be errors in the bids. The one sure thing that these errors will produce will be higher true costs through inefficient choices in the ultimate dispatch. There is no available evidence that the result would be lower prices. There are studies that suggest that both costs and prices would be higher.[26]

This general observation applied to any commodity auction applies with special force to something as complicated as the bids for a security-constrained economic dispatch. We saw what could happen in such a market when California operated fully separate energy, reserve and ancillary services markets.[27] In effect, this was an approximate prototype of a full pay-as-bid market. It was a stunning failure, the first in a line of special California problems. To cite another complication, consider the problems of transmission congestion management if everyone is bidding to make sure that the bid is close to the market-clearing price. For example, in PJM the presence of transmission congestion can change the market value of generation by an order of magnitude. Every generator would be compelled to consider the likelihood of transmission congestion in each interval, and change its bids accordingly. This embrace of a pay-as-bid rule would be a nightmare for the system operator and the competitive bidder, but a godsend for any generator who wished to cloak the exercise of market power.

Market Power and Shortages

High prices in the summer of 2000 arose because of a combination of factors. Faulty market rules created both inefficient dispatch and incentives for behavior that complicated market operations. Costs were up due to higher natural gas prices and tightening markets for emission allowances. Capacity was reduced because of the low availability of hydro power, a failure to invest in generating capacity in California, and increased congestion in the transmission system. Demand in areas not exposed to market prices grew at a rate that surprised most observers. On these points there is no dispute. In addition, there are those who argue that the high prices were exacerbated by the exercise of market power.

The need to fix badly flawed markets should be beyond dispute after the evidence of the failed experiment in California. The impacts of increased production costs and shortages are

---

[25] Federal Energy Regulatory Commission, "Staff Report to the Federal Energy Regulatory Commission on Western Markets and the Causes of the Summer 2000 Price Abnormalities," Part 1 of Staff Report on U.S. Bulk Power Markets, November 1, 2000, p. 5-15.

[26] John Bower and Derek W. Bunn, "Model-Based Comparisons of Pool and Bilateral Markets for Electricity," Energy Journal, Vol. 21, No. 3, pp. 1-29.

[27] Scott M. Harvey and William W. Hogan, "Issues in the Analysis of Market Power in California," October 27, 2000. (available at ksgwww.harvard.edu/people/whogan).

easy to understand, if not pleasant to endure. Markets respond to scarcity by increasing prices, and the increase in price creates the incentives for adjustments in supply and demand. Were it not for the large wealth transfer, the analysis of the proper response to scarcity would lead to the uncontroversial conclusion to let the market work.

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

Long-Term Market Reforms

The list of necessary reforms for the California market is long, and the difficulty of identifying and fixing all of the problems has been exacerbated by repeated *ad hoc* reforms that have dismissed theoretically sound and proven design principles. A transition will be necessary, but it must be guided by a set of principles that are consistent with a workable, efficient, and sustainable market. The necessary principles have been articulated in a number of different forms and forums.[28] Here we restate and summarize the key principles and their rationale.

1. *The ISO must operate, and provide open access to, short-run markets to maintain short-run reliability and to provide a foundation for a workable market.*

These short-run markets include, at a minimum, the real-time balancing market associated with the real-time dispatch, along with associated ancillary service markets – for regulation and operating reserves -- necessary to maintain reliability. A bid-based real-time dispatch is the means by which the ISO provides a real-time balancing and spot market, maintains system balance, and provides economic redispatch to manage congestion.

The FERC should reject the California restrictions on economic, least-cost dispatch for energy and ancillary services and refrain from imposing further restrictions or penalties on those who use the ISO's real-time market. The real-time market should be allowed to become an open, efficient spot market available to all market participants. To the extent that the ISO tends to have insufficient resources available to meet real-time loads, it should offer a unit commitment service to obtain those resources without restricting market choice.

2. *An ISO should be allowed to operate integrated short-run forward markets for energy and transmission.*

[28] For example, see the 17 design recommendations submitted to the Commission by San Diego Gas and Electric Company. "Comments of San Diego Gas and Electric Company on Order Proposing Remedies for California Wholesale Electric Markets, Attachment A," filed November 22, 2000.

Currently, the ISO is prohibited from operating integrated day-ahead forward markets for energy, even though it is charged with operating forward markets for transmission. However, the markets for energy and transmission cannot be separated without creating serious coordination problems that lead to inconsistent pricing and gaming between the markets. These inconsistencies can also lead to infeasible schedules that are accepted in the forward market but which force the ISO to redispatch in real time.

ISO-operated day-ahead and hour-ahead markets can provide useful options to market participants, allowing them to lock in energy and transmission (congestion) prices in advance of real-time. They also provide a mechanism for parties to exchange their transmission rights; that is, to settle their existing transmissions rights and gain new entitlements that match their scheduled transactions.

3. *An ISO should use locational marginal pricing to price and settle all purchases and sales of energy in its forward and real-time markets and to define comparable congestion (transmission usage) charges for bilateral transactions between locations.*

Several months before the California ISO and the FERC became preoccupied with the high prices produced by the California market, the FERC had already found the ISO's congestion management system to be "fundamentally flawed" and in need of comprehensive reform. Because the congestion management system implicates many other aspects of the overall market design, the ISO management's process for congestion management reform eventually grew into a comprehensive market redesign process. However, the most fundamental reform needed by the market design and the congestion management process is to get the prices right. The California zonal system is fundamental flawed because it cannot get the prices right. It is time for the California market to solve this fundamental problem by moving to nodal locational marginal pricing.

4. *An ISO should offer tradable point-to-point financial transmission rights that allow market participants to hedge the locational differences in energy prices.*

Once market participants are exposed to locational price differences and point-to-point congestion charges, they will need tradable transmission rights that allow them to hedge these locational differences and congestion charges in order to obtain *ex ante* price certainty for their transactions. Point-to-point FTRs will be necessary to support a nodal locational market price system.

The current FTRs are not point-to-point but are rather defined across specific inter-zonal interfaces. With the addition of at least eight more zones (LRAs) within California, the existing FTRs would have become increasingly unworkable. The existing FTRs are essentially a form of financial flowgate rights, and the addition of new zones would force market participants to struggle with the need to obtain multiple flowgate rights for each transaction, given the loops within and around the California grid. The ISO and stakeholders were only beginning to recognize the problems of changing distribution factors last Spring when they were overwhelmed by responding to the high price conditions. No clear solution to this problem has been proposed.

5. *An ISO should simultaneously optimize its ancillary service markets and energy markets.*

Experience in New England and California have now amply demonstrated that the short-run markets for regulation and operating reserves must be fully coordinated with the short-run markets for energy. Ideally, these markets should be simultaneously optimized and their pricing rules made consistent. This will ensure that generators receive efficient market-clearing prices in each market and are neither forced nor encouraged to guess at which market would be the more profitable venue. By optimizing these markets simultaneously, the ISO will ensure that the mix of resources chosen for energy and ancillary services will be the lowest overall cost, given the available bids. By using consistent pricing, generators will be assured that their cost recovery and potential for profits will not be adversely affected whether they are chosen to provide energy, provide regulation or spin, or withheld to provide reserves. If generators are paid consistent market-clearing prices in each market, they will not have to guess the market price or risk bidding mistakes. Instead, generators will have an incentive to bid their marginal costs.

6. *The ISO should collaborate in rapidly expanding the capability to include demand side response for energy and ancillary services.*

The least controversial reform of market design would be to implement all the changes needed to allow for demand side response in the face of higher prices. This should include changes both in the wholesale market mechanisms to allow for demand side bids in the day-ahead markets and, for properly metered and controllable loads, in the real-time market. In addition, retail rate designs under the control of the California Public Utilities Commission (CPUC) should be such that customers who choose can see the wholesale price and respond to higher prices by reducing their demands. Prices for usage should be based on the market-clearing level. Retail prices in California that are below the cost of fuel, subsidize electricity consumption in California and raise both electricity and gas prices throughout the west. Any rebates should be in terms of reduced connection costs or in some other manner to break the link between average and marginal rates.

Slightly more controversial, but equally important, would be to introduce the same type of demand response for reserves and ancillary services. Not all reserves are equally valuable, and there has always been some tradeoff between reliability and cost. The traditional procedures that embodied this fact have been replaced by rigid requirements in the new market that have the effect of forcing prices to very high levels, much higher than the reserves or the energy are really worth. The FERC has already addressed this issue in principle in the context of recent proceedings regarding the Northeast ISOs.[29] The same arguments apply to California. The CAISO should eliminate absolute reserve requirements in excess of the largest contingency and implement a demand curve, reflecting reserve shortages in day-ahead and real-time prices.

---

[29] For ISO New England, see for example, Federal Energy Regulatory Commission, "Order Conditionally Approving Congestion Management and Multi-Settlement Systems," Docket No. EL00-62-000, June 28, 2000.

NEW ZEALAND REFORMS

The combined effects of market successes in the United States, as well as the prominent failure in California, suggest a number of lessons for consideration in New Zealand. In many ways, the New Zealand market design has been at the forefront of best practice. The review and restructuring effort in New Zealand provides an opportunity to extend that design.[30] However, experience suggests that every discussion to rethink market design risks breaking what is not broken. The task is to recognize what is essential, what works well, and what experience elsewhere suggests would be a mistake that should not be repeated.

Market Design

The New Zealand electricity market provides fundamental design elements needed to support competition in generation and supply. A key feature of any such market is the use of a coordinated spot market to handle balancing, transmission usage and security requirements. The New Zealand spot market includes a bid-based, security-constrained, economic dispatch for real-time decisions, which is the most natural and effective means of providing an array of services essential for successful operation of the market and reliable operation of the electricity system. The bids summarize the preferences of the market participants and ensure that the final dispatch choices respect those preferences. The security constraints preserve the conditions needed to ensure reliable operations. The principles of economic dispatch define both the traditional engineering practice and the results of a competitive equilibrium. In this regard, the New Zealand model for real-time operations is aligned with the best international practice for a competitive electricity market.[31]

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

Efficient Pricing

Efficient pricing is a central feature of a competitive electricity market. It is essential if the benefits of a competitive market are to flow through to customers and other market participants. Pricing that is inefficient, on the other hand, will fail to signal and encourage

---

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

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

appropriate levels of consumption and supply or the appropriate levels and locations of new generation and transmission investments.

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

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

As a matter of principle, these locational prices are simply the market-clearing prices based on all the bids and the details of the requirements of network operations. Furthermore, for any given economic dispatch, it is an easy matter to determine these prices based on the bids and the system conditions. These locational prices are in use today as an integral part of the New Zealand market design.

In addition to defining the market-clearing price at each location, these locational prices provide an immediate and simple answer to the otherwise intractable question as to the appropriate marginal cost or market-clearing price of transmission use. The electric network is complicated, with the power flow dictated by the laws of physics and many system constraints. Tracing the details of transmission flow has proven to be a blind alley that has frustrated attempts to define workable methods of transmission pricing.[32] But the locational pricing approach that accompanies the coordinated spot market provides an immediate simplification of this difficult problem. In particular, transmission of a megawatt between two locations is physically equivalent to sale at the source and purchase at the destination. In equilibrium, therefore, the market-clearing price determined by the marginal cost of transmission must be the same as the net price for the combined purchase and sale transaction. In other words, the price of transmission between two locations must be just the difference in the locational prices of energy.[33]

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

---

[32] William W. Hogan, "Flowgate Rights and Wrongs," Center for Business and Government, Harvard University, August 2000.

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

sales through the spot market should be at the locational prices. Any scheduled transmission between locations should be charged at the difference in the locational prices. This basic model is operational in New Zealand as well as the Pennsylvania-New Jersey-Maryland Interconnection (PJM) in the United States, New York, Argentina, Chile, and so on.

The details of locational pricing differ slightly for congestion and losses. Some regions, such as PJM, have implemented only congestion pricing, but intend to include losses with planned modifications of their software. Others like New York are similar to New Zealand in representing both losses and congestion. Given a reference location for the cost of generation, the price at other locations can be decomposed into the marginal cost of congestion and the marginal cost of losses. The congestion cost arises when the transmission system is constrained and it is necessary to redispatch by reducing output for some cheaper plants and increasing output at some more expensive plants. The difference in cost is the congestion charge. If the system is not constrained, then there is no need for redispatch and the congestion charge would be zero. Except for losses, therefore, in the absence of congestion the spot electricity price would be the same throughout the gird, an intuitive outcome. In the presence of constraints, the contribution to congestion can be either positive or negative at a location depending on whether or not increased load at that location adds to or subtracts from the flow over constrained transmission elements.

The effect of losses depends on the location of load and generation. Increasing the load at a location, and balancing the increase with increased production at the reference location, will change the flows throughout the system. The flow along transmission lines induces resistive losses. In some locations, increasing the load will increase the total system losses, and the marginal loss component of prices would be this change in total losses times the reference price. In other locations, there could be a reduction in losses leading to a negative contribution to the locational price.

For both losses and congestion, the contribution for any line or location can be unpredictable or even undefined. However, for both losses and congestion the total contribution is always positive. The aggregate congestion payments are zero only if the system is unconstrained, and positive when the system is constrained. The aggregate loss payments are always positive. Therefore, these net locational payments produce a transmission rental that captures the total congestion cost and the difference between marginal and average losses.[34]

An issue then arises as to the best use of these transmission rentals. In the present New Zealand system, these rentals are used to reduce the transmission charges for access to the grid. The basic logic is that the payment should be divorced from the marginal usage decisions, in order to preserve the incentives of efficient pricing. Further, it is intuitive that the proper recipients of the rentals should be those who are paying the transmission access charges to cover the fixed costs of the grid. This logic is consistent as far as it goes. However, as we shall see, there is another and superior use of these rentals in funding long-term transmission rights.

---

[34] William W. Hogan, "Regional Transmission Organizations: Millennium Order on Designing Market Institutions for Electric Network Systems," Center for Business and Government, Harvard University, May 2000.

Long-Term Transmission Rights

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

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

Traditionally, and as is seen in many other markets, the notion of a long-term contract carries with it the assumption that customers and generators can make an agreement to trade a certain amount of power at a certain price. The implicit assumption is that a specific generator will run to satisfy the demand of a specific customer. To the extent that the customer's needs change, the customer might sell the contract in a secondary market, and so too for the generator. Efficient 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. 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 and no losses. In this circumstance, it is possible to treat all production and consumption as at the same location. Here the natural arrangement is to contract for differences against the equilibrium price in the market. A customer and a generator agree on an average price for a fixed quantity, say 100 MW at five cents. On the half hour, if the spot price is six cents, the customer buys power from the 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 pool price at three cents, the customer pays three cents to the pool, 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. This then is the familiar "contract for differences." It allows for long-term contracts without direct administration by the system operator.

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 system operator other than for the continuing short-run market transactions. But through the interaction with system operator and the coordinated spot market, the situation is even better than with a long-run contract between a specific generator and a specific customer. For now if the customer demand is above or below 100 MW, there is a ready and an automatic spot market where extra power is purchased or sold at the spot price. Similarly for the generator, there is an automatic spot market for surplus power or backup supplies without the cost and problems of a large number of repeated short-run bilateral negotiations with other generators. And if the customer really consumes 100 MW, and the generator really produces the 100 MW, the economics guarantee that the average price is still 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 need have no knowledge of the terms.

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

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

The convenient solution to both problems – providing a price hedge against locational congestion differentials and removing the adverse incentive for the system operator – is to redistribute the congestion and net loss revenue through a system of long-run financial transmission rights (FTR) 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 an FTR that provides compensation for differences in prices, in this case for differences in the congestion and marginal loss costs between different locations across the network.

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

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

In the presence of losses, full implementation of FTRs requires at least some of the rights to be unbalanced, with different quantities at the source and destination. In principle, any FTR would be represented as a combination of a balanced FTR and a spot FTR for input or output at either location. The eventual requirement is not that each individual FTR matches the corresponding losses, but that the aggregate of all FTRs be simultaneously feasible including losses.

When only the single generator and customer are involved, this sequence of exchanges 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 allocated transmission flows is feasible. However, the net result would be the same. Short-run incentives at the margin follow the incentives of short-run opportunity costs, and long-run contracts operate to provide price hedges against specific quantities. The system operator coordinates the short-run market to provide economic dispatch. The settlement system collects and pays according to the short-run marginal price at each location, and then distributes the transmission rentals to the holders of FTRs. 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 FTRs 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 FTRs awarded to customers, the aggregate congestion payments received through the spot market will fund the payment obligations under the FTRs. Still, the transmission prices paid and received will be highly variable and load dependent. Only the system operator will have the necessary information to determine these changing prices, but the information will be readily available embedded in all the spot market locational prices. The FTRs define payment obligations that guarantee protection from changes in the transmission rentals.

Defining transmission rights in this way as financial contracts exploits a simplification that is available with a coordinated spot market and efficient locational pricing. However, nearly everyone first confronted with the problem of transmission usage instinctively looks for a system of physical allocations of tradable rights that could be used in a decentralized market. In this regard, there is some advantage in drawing the conceptual connection between physical rights and FTRs.

The FTR can be recognized as equivalent to an advantageous form of "physical" transmission right. 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 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 such physical transmission rights that parcel up the transmission grid. However, were it possible then the rights would either be used by the owner to transmit power or sold to others at the market-clearing price. In either case, the net result would be the same as with ownership of the FTR. In a competitive electricity market with a bid-based economic dispatch, therefore, FTRs are equivalent to just such perfectly tradable transmission rights. Hence we can describe FTRs either as financial contracts for transmission rents or as perfectly tradable physical transmission rights.

The interpretation of FTRs as equivalent to perfectly tradable physical rights makes it self-evident that FTRs can only be offered by the system operator. Only the system operator can deal with the complex interactions among transmission flows in the physical sense, or in the equally complex interaction of transmission rentals in the financial sense. Furthermore, just as the system operator would have to respect the physical capacity of the system in allocating physical rights, so too must the system operator respect the physical capacity in allocating FTRs. As long as the FTRs allocated would be simultaneously feasible on the existing grid, the payment obligations under the FTRs at the current spot prices will be no more than the transmission rentals collected through the spot market settlements.

This allocation of transmission congestion and loss rentals is not the same thing as assigning the value of the rentals to the system operator. The rentals pass through the system operator, the only entity with the information necessary to resolve the complex interaction and allocation process for electric gird usage. But the system operator acts only as an agent for the market, not as a party to the transaction.

If the FTRs have been fully allocated, then the system operator will be simply a conduit for the distribution of the transmission rentals. The revenues from the spot market would fund the payments under the FTRs, so there is no financial risk for the system operator. The operator would continue to have no incentives to increase transmission rentals: any increase in rental payments would flow only to the holders of the FTRs. And through a combination of generation contracts and FTRs, participants in the electricity market can arrange price hedges that could provide the economic equivalent of a long-term contract for specific power delivered to a specific customer.

In practice, FTRs have been defined as "obligations" in the sense that the price difference can be negative, in which case the payment is from the holder of the FTR to the settlements system.[35] In principle, the FTR could be defined as either an option (positive payments only) or an obligation (both positive and negative payments).[36] The technical differences relate to the treatment of counterflow and the ability of some transactions to relieve transmission constraints. In order to have the incentive to provide counterflow, obligations are required. The use of options complicates determination of the simultaneous feasibility of the FTRs.

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 sometime 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 aggregation of nodes. 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.

In the case of FTR obligations, the same decomposition applies to match with the decomposition of transmission pricing. However, this decomposition is not available for FTR options. An FTR option between two locations is not the same as two options, to and from the hub.

---

35 The obligation definition is applied in PJM and New York.

36 The proposal for New England includes a directive from the Federal Energy Regulatory Commission to include FTR options. For a further discussion see, William W. Hogan, "Flowgate Rights and Wrongs," Center for Business and Government, Harvard University, August 2000.

The initial allocation of FTRs could be according to any of a number of rules. In PJM most FTRs are allocated by formula to those who are paying the transmission charges for access to the grid, with the remainder being auctioned. In New York, there is an auction to award the FTRs to the highest bidders, with the revenue from the auction used by formula to reduce the transmission charges. Other things being equal, there are advantages to the auction approach for improving the liquidity of the FTR market. Furthermore, the auction ensures that the full capacity of the grid is available for allocation without any bias in the choice of recipients other than their willingness to pay in a competitive bidding process. But in either approach the actual transmission rentals return to those who pay the transmission charges, either through direct ownership of the FTRs or through the auction revenues, which can be expected to be equal to the expected transmission rentals.

Long-Term 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 hook-up. 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 real-time 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 the local nodal 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 spot price.

If either party expects significant transmission congestion or losses, then an FTR would allow the parties to hedge their price risk for any case that would otherwise benefit from point-to-point physical rights were they available. If FTRs 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 FTRs. In this way, the availability of FTRs converts the short-term transmission price into a long-term investment signal. The system operator need participate in the process only to verify that the newly created FTRs would be feasible and consistent with the obligation to preserve any existing set of FTRs 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 FTRs 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 FTRs.

This fortuitous 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.[37] In this case, however, the task should be simplified (if not solved) by the ability to simultaneously allocate the benefits in the form of a share of the FTRs to those that invest. The market could take care of many, perhaps most, investments, and the regulatory option would be easier to implement when needed.[38]

A related problem could appear in the circumstances where the pattern of transmission 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 equally true, of course, for physical rights were they possible, as well as for the FTRs. 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 again would not follow without some regulatory entity that can mandate payment of the costs.

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 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 and marginal loss costs, versus avoiding the sunk costs while using the system and paying the continuing cost of congestion and losses, 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 transmission access charges would be the functional equivalent of the contract payments for new investment.

---

[37] 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.

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

This same basic system has been in operation for more than two years under the PJM Interconnection, where the financial transmission rights are labeled as fixed transmission rights (FTR).[39] 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).[40] In the proposed New England ISO structure, the term of art is a financial congestion right (FCR).[41] 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.

These successes did not come easily. In PJM and New England, the initial market designs proposed by market participants contained features that did not work well or at all.[42] As with California, the FERC initially accepted the voluntary agreements of the market participants, but later unhappy experience forced the regulators to push for better systems that supported the public interest in efficient markets. This role for the regulator as one of the few groups charged with pursuit of the public interest is important in considering the approach to governance in New Zealand.

Governance

The Government of New Zealand set down principles for reform of the electricity market and development of new regulatory arrangements.[43] The principles are excellent in defining the objectives for an electricity market with the emphasis on the importance of the broad public interest. There is a tension, however, between a bias to rely on voluntary industry solutions and the possible need for mandatory regulatory prescriptions. Here the experience in the United States, with its practice of formal regulatory oversight, can illustrate the importance of governance explicitly charged with the public interest as an essential ingredient in achieving the right balance.

The New Zealand "Government favours industry solutions where possible, but is prepared to use regulatory solutions where necessary." [44] At a high level of generality, it is easy

---

[39] 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.

[40] 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.

[41] 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.

[42] 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.

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

[44] Pete Hodgson, Minister of Energy, "Government Policy Statement: Further Development of New Zealand's

to agree with the bias that, other things being equal, voluntary arrangements are better than mandatory rules. Furthermore, other things being equal, we should leave the burden of proof with those who wish to replace a voluntary arrangement with a mandatory structure.

There are many reasons for this bias, not just promoting economic efficiency, and these have much to do with the political philosophy of democracy. But there is a closely related line of argument that might be implied by the bias, and this is relevant in designing market institutions. This alternative motivation would be the view that voluntary arrangements naturally tend to work out of inefficient solutions, because if a better arrangement were possible the participants would voluntarily choose to adopt the preferred solution.

A leading proponent of this argument for institutional efficiency was the Nobel Prize winning economist Douglass North. What is less well known is that when faced with overwhelming evidence to the contrary, North later recanted and turned to argue that path dependence and other institutional failures could lead to voluntary arrangements that were inefficient but effectively permanent.

> "If political and economic markets were efficient (i.e., there were zero transaction costs) then the choices would always be efficient. That is the actors would always possess the true models or if they initially possessed incorrect models the information feedback would correct them. But that version of the rational actor model has simply led us astray."[45]

Other things are seldom equal. The choices of stakeholders can be inefficient for many reasons, and the compromises among stakeholders that result in the least common denominator may not serve the public interest. In this case, when the problem is more like designing a bridge than dividing a pie, government can play a critical role in designing the institutions. Of course, government can fail just as can markets. This implies that we have to consider the tradeoff between mandatory government rules and voluntary market solutions.

We may seek to limit the scope of government, but require that mandatory rules be well designed. Hence, when faced with some activity that is voluntary and consistent with normal patterns of commerce, government need not be involved and should defer to the voluntary choices in the market. However, when there is something that must be done with direct or indirect appeal to the power of government to compel action (and hence make mandatory rules) there is no reason to presume that the voluntary agreements among stakeholders should be preferred or accepted. In other words, the voluntary agreements among some about how they will compel others should not be presumed to be superior in the implicit pursuit of efficiency, as compared to a design that is explicit about the support of public purposes.

The United States experience in electricity design has been that voluntary agreements on the core market design issues have often been flawed, and the regulatory authority of government has been essential in crafting workable institutions. The FERC has been alternatively deferential and prescriptive. Deference to flawed markets in the early voluntary choices of stakeholders has been replaced with more prescriptive approaches, such as in the

---

Electricity Industry," Wellington, New Zealand, December 2000, p. 1.

[45] Douglass C. North, Institutions, Institutional Change and Economic Performance, Cambridge University Press, 1990, p. 8.

cases earlier in PJM and New England, and later in California. At the end of 2000 this issue was still not fully resolved in the United States, as witnessed by the slow progress during the year in actually implementing the prescriptions of the Millennium Order on regional transmission organizations. However, the lesson is there that deference can be dangerous; there is no magic in so-called voluntary agreements to invoke the mandatory authority of government. It is one thing to allow market participants to arrange their own trades within the rules, it is quite another thing to allow the market participants to choose the rules.

Most recently, consider the case of California. The stakeholders voluntarily agreed that they preferred decentralized trading to centralized coordination of the spot market. They then used this voluntary agreement among the majority to impose mandatory rules on the California ISO that would compel everyone to pursue decentralized trading and prevent them from having the option of a coordinated spot market. (This is embodied in California's market separation principle.) This rule has had many perverse effects, and is demonstrably inefficient. However, it was in the interest of many of the stakeholders and has proved difficult to change. There were no truly voluntary features in any of this, it was simply a perverse distortion of the logic which compounded the problems in this failed market. Fortunately, the regulators have now ordered these practices abolished, and we may see more sensible rules in the future.

By contrast, the independent Automated Power Exchange (APX) arose in California to provide another decentralized trading alternative.[46] This APX has no powers to compel anybody to do anything. And there is and should be no concern about its governance other than the normal rules of commerce.

The distinction, therefore, is to defer to voluntary arrangements when no powers of government are invoked. But when the powers of government will be employed, directly or indirectly, the responsibilities of government come hand-in-hand. By this criterion, scheduling, dispatch, access pricing, congestion pricing, settlements, transmission expansion, and so on would fall under the rubric of government mandates for efficient institutions that operate in the broader public interest. These are all activities where there will have to be mandatory rules, and if participants don't like the rules they will not be able to conduct their business elsewhere. Furthermore, the rules will interact, and it will be no easy matter to make sure that separate governing arrangements can ensure their consistency.

By contrast, long-term trading and energy contracting do not have the same features or need to impose any rules other than the normal features of commercial law. Hence, governance of these activities does not impose any special obligations that are not familiar from other industries. There is no need to invoke the power of government as applied to deal with the special characteristics of electricity systems. Here there can be a bias in favor of voluntary arrangements.

Once the power of government is invoked, however, there might even be an opposite bias. Namely, voluntary agreements among the majority should be viewed with skepticism to avoid the tyranny of the majority; or to avoid descent to the least common denominator. And stakeholder processes will be both inefficient and difficult to affect. The time to emphasize democracy is at the stage of legislation to create the authority and establish the principles. But implementation is better left to independent agencies that can exercise their good judgment in

[46] See www.apx.com.

designing the detailed rules. We don't want to go to Congress to modify the rules of FERC. And FERC does not view itself as a stakeholder board. However, the FERC does answer to Congress. This is a balance that is difficult, but seems right.

The governance model for the Financial Accounting Standards Board (FASB) is another instructive example. The stakeholders elect the Board members, but the Board members must be independent and with no existing ties to any of the stakeholders. They serve for a long term, during which they make rulings about accepted accounting standards that are effectively mandatory. Their charge is to safeguard and promote the public interest in accounting reports and procedures, not just to serve the interests of the interested market participants. The insulation of the Board is not perfect but is viewed as absolutely essential, and the process would collapse if the stakeholder were directly involved in setting the rules.

The challenge for government and governance is clear and substantial. The electricity industry is complex, and the process of separating the components to operate in a competitive market imposes a greater burden to make sure that the pieces fit together. We know from experience that the design of market institutions can be critical. There is no reason to expect that good institutions will appear without the visible hand of government setting the rules for what is essential in order to provide the right framework for the invisible hand of the market to do its magic.

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. Government and governance mechanisms should keep these principles at the forefront in pursuit of the public interest.