10/1-26-96 JS(1)

LBL-37508
UC-1320

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

ENERGY & ENVIRONMENT DIVISION

**Organization of Bulk Power Markets:
A Concept Paper**

E. Kahn and S. Stoft

December 1995

ENERGY
AND ENVIRONMENT
DIVISION

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

Available to DOE and DOE Contractors
from the Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, VA 22161

LBL-37508
UC-1320

Organization of Bulk Power Markets: A Concept Paper

Edward Kahn and Steven Stoft

Energy & Environment Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

December, 1995

The work described in this study was funded by the Assistant Secretary of Energy Efficiency and Renewable Energy, Office of Utility Technologies, Office of Energy Management Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Contents

Figures

Tables

Acknowledgments

We would like to express our appreciation to the following people who provided valuable review comments on an earlier draft:

Theresa Flaim
Niagara Mohawk Power Corporation

Charles Goldman
Lawrence Berkeley National Laboratory

Eric Hirst
Oak Ridge National Laboratory

Ben Hobbs
Johns Hopkins University

William Hogan
Harvard University

Val Jensen
U.S. Department of Energy

Kevin Porter
National Renewable Energy Laboratory

Kathy Treleven and Jim Davidson
Pacific Gas and Electric Company

The work described in this study was funded by the Assistant Secretary of Energy Efficiency and Renewable Energy, Office of Utility Technologies, Office of Energy Management Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Acronyms and Abbreviations

ACE	Area control error
ADOE	Alberta Department of Energy
BRPU	Biennial Resource Plan Update
CCGT	Combined cycle gas turbine
COB	California-Oregon border
CPUC	California Public Utilities Commission
FERC	Federal Energy Regulatory Commission
GAPP	General Agreement on Parallel Paths
IPP	Independent Power Producer
ISO	Independent system operator
NERC	National Electric Reliability Council
NUG	Non-utility generation
NYMEX	New York Mercantile Exchange
NYPP	New York Power Pool
OFFER	Office of Electricity Regulation
PJM	Pennsylvania New Jersey Maryland Interconnection
PUC	Public Utility Commission
PURPA	Public Utilities Regulatory Policy Act
PSI	Public Service of Indiana
QF	Qualifying Facilities
REE	Red Electrica de Espana
RTG	Regional Transmission Group
SCD	Security constrained dispatch
SDG&E	San Diego Gas and Electric Company
WSCC	Western System Coordinating Council

Executive Summary

Introduction

The electricity industry in the U.S. today is at a crossroads. The restructuring debate going on in most regions has made it clear that the traditional model of vertically integrated firms serving defined franchise areas and regulated by state commissions may not be the pattern for the future. The demands of large customers seeking direct access to power markets, the entry of new participants, and proposed reforms of the regulatory process all signify a momentum for fundamental change in the organization of the industry. This paper addresses electricity restructuring from the perspective of bulk power markets. We focus attention on the organization of electricity trade and the various ways it has been and might be conducted.

Our approach concentrates on conceptual models and empirical case studies, not on specific proposals made by particular utilities or commissions. We review a large literature in economics and power system engineering that is relevant to the major questions. Our objective is to provide conceptual background to industry participants, e.g. utility staff, regulatory staff, new entrants, who are working on specific proposals. While we formulate many questions, we do not provide definitive answers on most issues. We attempt to put the industry restructuring dialogue in a neutral setting, translating the language of economists for engineers and vice versa. Towards this end we begin with a review of the basic economic institutions in the U.S. bulk power markets and a summary of the engineering practices that dominate trade today.

International Experience with Restructuring

Electricity restructuring is a worldwide phenomenon. Because of this, the experiences of other countries may provide some useful perspective for the U.S. debate. We survey this experience.

The restructuring process in electricity can be divided into three general elements. First, vertically integrated firms are reorganized to separate generation from the transmission and distribution assets. We call this functional restructuring. It includes both divestiture of generation and a re-orientation of the transmission function to facilitate increased competition. Because the organization of the competitive process can be so varied, we refer to generation competition as a separate stage. Finally, there has frequently been a privatization element associated with electricity restructuring. In most countries outside of the U.S., electricity industry reform originates from a situation of public ownership.

Table ES-1 summarizes the restructuring process in a number of countries. It shows both the varying extent of implementation and the substantially different starting point of the U.S. Because of private ownership in the U.S., functional restructuring is linked to stranded cost recovery. Once the possibility of expanded competition arises, private investors need assurance of cost recovery for assets that would become economically obsolete, if they are going to cooperate with such a transition. The cases of Spain and Alberta (Canada), where private ownership has a long history, involve recent functional restructuring but limited competition to date.

Table ES-1. How You Proceed Depends Upon Where You Start

Country	Functional Restructuring	Generation Competition	Privatization
UK, Chile, Argentina, Victoria (Australia)	———	———	———→
Norway	———	———→	
NSW (Australia) **New Zealand**	———	→ ?	
Ontario (Canada)	? →		
U.S.		←———	———
Spain, Alberta (Canada)	———→		←———

International restructuring experience offers few major lessons for the US, because none of these situations involves both the pervasive private ownership and the highly fractionated nature of the US electricity industry. The importance of private ownership is that it makes functional restructuring more complex than where government enterprises are involved exclusively. Where restructuring involves writing down the value of uneconomic assets, governments accomplish this is more easily than if private investors must be compensated. The large number of relatively small utilities in the US power network means that the formation of pools would also involve complicated financial compensation procedures. In Chile, Argentina and the UK, a centralized pool operates the wholesale market, with all trade going through the pool. These pooling institutions were formed from pre-existing national utilities. A bilateral trading regime dominates the Norwegian market, but it depends upon an uncongested transmission network, and is facilitated by a pre-existing framework of pooling and marketing institutions. Australian interstate trading experiments promise greater market decentralization than the pool cases, made simpler by the presence of few market centers.

Constraints may limit electricity restructuring. Where technology involves environmental or economic externalities, such as hydro or nuclear generation, there is likely to be public ownership, because private markets cannot easily internalize all of the associated externalities. These cases may pose potential barriers to privatization. In Norway, a competitive regime has not included privatization of the generation, 98% of which is hydro. Other important constraints include the nature of local fuel markets and the strength of the transmission network. Competitive electric markets are facilitated by competitive fuel markets. Where coal mines are vertically integrated with generation plants, the competitive situation may be complicated. This situation arises in Spain, Australia and Alberta. In many countries, natural gas plays an important role facilitating competition by lowering the barriers to entry. Gas-fired combined cycle plants are relatively simple to build and operate, have low capital costs and very high thermal efficiency. Where a well developed natural gas infrastructure exists, competition will be more robust. The transmission network is the vehicle through which electricity competition occurs. The stronger this network is, the smaller the limitations of network congestion. When the network is congested, markets are geographically separated, limiting competition.

Market Power Concerns Are Important

We can distinguish three sources of market power in the electricity industry: market concentration in generation, vertical integration that may limit access for competitors, and the ability to block transmission pathways.

Market concentration is the typical source of market power and will probably prove to be the most decisive factor in bulk power markets as well. We show that forward contract markets tend to ameliorate the effects of market concentration, while network congestion increases the level of market concentration. Our analysis also suggests that a limited amount of market power derived from concentration is found to play a helpful role in maintaining system reliability.

Perhaps the central question concerning the organization of U.S. bulk power markets is whether vertical re-organization of the entire industry will be required. We analyze various forms of market power engendered by collusion between the DistCo and GenCo halves of vertically integrated firms. Specifically, we discuss the potential for this collusion to inhibit entry by other GenCos and the possibility that an Independent System Operator (ISO) can control this type of behavior.

A particular kind of access limitation that is peculiar to electricity is the possibility for strategically located Gencos to block transmission access for other suppliers. We illustrate how this can be achieved in an electric network and observe that such practices seem to have occurred in other countries.

Ensuring "Open Access" Will Likely Require an Independent System Operator

Achieving open access in electricity markets is the objective of the FERC's Mega-NOPR. We argue that FERC's open access objectives simply cannot be met successfully without the use of an ISO, and that even with one, the pre-determined transmission prices envisioned by FERC impede the implementation of an efficient access rule.

Both major models of electricity competition, the bilateral and the spot-market approaches, rely on an Independent System Operator (ISO) to provide non-discriminatory access. FERC attempts to achieve non-discriminatory access without the benefit of an ISO. Open access must solve the problem of excess demand for transmission to preserve reliability, and the FERC's Mega-NOPR allows this only through a predetermined tariff and non-price rationing. Because of the complexities of transmission, non-price rationing cannot be accomplished in a non-discriminatory fashion by an interested party. This leaves the predetermined tariff as the only tool for preventing excess demand. But for this to be almost perfectly effective, as is required by reliability, the tariff must be so high as to prevent even moderately efficient use of the grid.

Summary of the Bilateral vs. Spot Market Debate

A few broad principles have emerged from recent debates and experience both in the US and other markets. These include:

(1) a general agreement that price transparency is desirable,
(2) a broad (but not universal) consensus that an independent system operator (ISO) is necessary to facilitate increased trade,
(3) an increased skepticism about vertical economies, and
(4) a consensus that market distortions should be minimized.

The two main competing market models, spot market pools versus bilateral trade, would implement these principles differently. Table ES-2 summarizes the differences.

Table ES-2. Market Model Summary

	Pool	Bilateral
ISO Function	Grid Merchant	Information Broker
Price Transparency	SRMC	Index
Vertical Economies	less important	more important
Contract Performance	Financial	Physical

The ISO function involves a broader range of actions in the Pool model than in the bilateral model. The reason is that the Pool ISO is dispatching the power system based on sellers' prices, whereas the bilateral ISO is an information broker who facilitates the trading decisions of others. The different conceptions of the ISO are reflected in all of the other market model attributes in Table ES-2.

Price transparency facilitates competition by making the value of power clear to participants, but the notion means different things in the two models, because price formation differs in each. When the Grid Merchant is the central clearinghouse, the resulting prices at any network node are the short-run marginal cost (SRMC) at that node. Where no congestion exists, there is effectively a single market clearing price for any given period of time. In the Pool model, price variance results from the time differentiation of SRMC, not from any variance at a given time. The bilateral trade model is more compatible with price indices, which are typically derived from market information with inherently longer time horizons than hourly bids from generators used to estimate SRMC in the Pool model. These indices are averages of many bilateral contract prices. The potential biases in price reporting will also differ in each model. In the Pool model, there is some arbitrariness in SRMC determination. In the bilateral model, sampling error may distort price indices.

The question of vertical economies is quite unsettled. Both competitive models inherently question the role of vertical economies. In neither case, however, is it clear that divestiture of generation will be required for unbiased functioning of bulk power trade. There appears to be less emphasis on vertical economies in the Pool model, if only because of the international precedents, where electricity restructuring along pool lines has been accompanied by vertical separation. The bilateral model also seems more consistent with a vertical structure, because it is closer to current U.S. industry structure and practice. The increasing occurrence of utility mergers may end up raising market power questions in either of these models. The traditional arguments for vertical economies in a monopoly structure may turn out to look like access barriers in a competitive model.

Finally, contract performance standards differ in the two models. Given that the Pool ISO is a Grid Merchant through whom all physical transactions clear, the only role for contracts is financial. Indeed, a physical performance standard, where seller must physically deliver to buyer, is incompatible with the Grid Merchant concept. In the bilateral model, on the other hand, physical performance is the essence of commercial relations. It embodies the mutual commitment of the parties to trade. Physical performance as the cornerstone of a bilateral trade market may impose some complexity on the ISO, but proponents argue that this is feasible.

Future Research

The FERC Mega-NOPR lays out the clear objective of achieving open access, but it is less persuasive on the means of achieving it. The major conceptual models of electricity competition are incompatible with the Mega-NOPR framework in a number of ways. The choices concerning the future organization of bulk power markets will be influenced by a number of factors. These include the magnitude of the transactions costs involved, the impact on reliability of increased competition, and the potential for abuse of market power. Of all these issues, it is probably market power which will be the most decisive. If market power problems are found to be excessive, it is likely that vertical separation will be required. Whether this turns out to be the case is the major structural uncertainty in bulk power markets in the U.S.

1 Introduction and Overview

There is a significant demand for increased electricity trade in the U.S. today. The principal signs of this demand are the recent appearances of electricity marketers, who hope to profit from trade opportunities, and the demand for retail access by large industrial consumers. The regulatory response to these market developments at the federal level began with the Energy Policy Act of 1992 and continues with Federal Energy Regulatory Commission (FERC) actions, the latest of which is the Mega-NOPR of March, 1995. At the state level, numerous initiatives to experiment with retail access have been initiated, the best publicized of which is the California Public Utilities Commission (CPUC) Blue Book proposal of April, 1994 (CPUC 1994). All of this activity, in one way or another, calls into question the current functioning of the bulk electricity market. Responding to the demand for increased trade implies major restructuring of the electricity industry. It is the purpose of this concept paper to address the bulk power market issues raised in the debate about the organization of the electricity industry.

Increased electricity trade means that the electricity industry will become more competitive, and hopefully more efficient. A large part of the controversy over these developments involves uncertainties about the potential magnitude of the efficiency gains, their origin, and the practical feasibility of achieving them. It is entirely possible that the pursuit of these gains may sacrifice other benefits of the current structure. The questions raised by restructuring, therefore, are quite fundamental. In this paper, we will attempt to formulate these questions as precisely as possible, and review the controversies that have been generated in the discussions to date. Many of the other important issues associated with restructuring, such as stranded cost compensation and the fate of social programs, lie outside the bounds of this discussion.

The paper is organized in the following fashion. Chapter 2 reviews the basic economic institutions in the electricity industry and formulates the problem of the optimal industry structure as a trade-off between competition and coordination. Electricity involves managing significant technical constraints; we summarize current engineering practices used to address these constraints. Chapter 3 addresses international experience with electricity restructuring. The forces driving restructuring are world wide, and there may be valuable lessons from the experience of other countries. Chapter 4 looks more closely at current practices in the U.S. and focuses on significant decentralizing forces that are operating in the existing bulk power market. These include the role of non-utility generation, developments in the direction of greater price transparency in the wholesale market, and assessments of the institutions that currently organize electricity trade. In Chapters 5 and 6, we analyze proposals to increase competition and assess the problems raised by a competitive organization of bulk power markets. Chapter 5 discusses various proposals to facilitate increased trade in bulk power markets. We contrast the market model organized around the concept of nodal spot prices (usually implemented in a pool), with various kinds of bilateral trading models. In Chapter

6 we address several key problems posed by the presence of market power. We examine both traditional issues in market concentration and those specific electricity issues that may mitigate or exacerbate market power. We also consider limits on access associated with vertical integration. We conclude that an independent system operator (ISO) will be necessary for an effective open access trading regime. Chapter 7 discusses transmission capacity expansion and presents several ways to handle this issue in a competitive setting, including the possibility of allowing private investment in new lines. Conclusions and suggestions for future research are presented in Chapter 8.

2 Economic Institutions and Engineering Practices: The Basics

This chapter reviews the basic elements of bulk power markets from the economic and engineering perspectives. First, we summarize the economic literature on U.S. institutions that organize the buying and selling of electricity. We focus on that element of the economic literature which addresses broad questions concerning the organization of electricity markets, rather than the substantially larger literature addressing particular sub-markets. We then describe the engineering practices used in the U.S. to produce and deliver electricity. This discussion will establish the general questions posed by the demand for increased electricity competition and trade, and characterize the technical constraints affecting the electricity marketplace.

2.1 Economic Institutions

Economists have looked closely at many particular aspects of the U.S. electric power industry, but there have been relatively few studies which directly address the optimal industry structure. We survey a small number of these more general studies to highlight the major issues that are posed by restructuring.

2.1.1 *Markets for Power*

The book, *Markets for Power*, by Joskow and Schmalensee (1983) is the *locus classicus* of electricity deregulation analysis. It was written before the wave of international experimentation with electricity restructuring began, but it nonetheless reflects many of the basic issues that neither theory nor experience have resolved about the optimal structure of the electricity industry. Rather than recapitulate the specific industry models that Joskow and Schmalensee (JS) use to frame their analysis, it is more useful to identify the main characteristics of the industry on which they focus, to outline the criteria they use to assess deregulation, and to review their predictions in light of subsequent developments.

According to JS, increasing competition and trade in the electricity industry poses a trade-off between the potential efficiency benefits (in both the short and long run), and the increased transactions costs and the potential abuse of market power. JS pay substantial attention to the transactions costs associated with achieving in a competitive setting those coordination economies, such as efficient investment and dispatch, that are currently accomplished through allocations that are internal to the vertically integrated utility. Because many complex contingencies must be taken into account in a contracting regime, there may be rigidities and inefficiencies that result from these arrangements. JS focus attention on the key role played by transmission in the integration of supply and demand in any organization of the industry. Given the externalities involved in electric networks, where actions at one point have effects elsewhere, coordination economies currently achieved may be difficult to duplicate in a competitive setting. Congestion in transmission networks may also confer substantial market power on generating companies in a less regulated electricity industry. JS make only the most approximate estimate of market concentration, but find that it is likely to be a concern.

Many of these points have been confirmed by subsequent experience. We address all of them in one way or another in this study. Since JS original analysis, perhaps the main changes in the climate of opinion concerning electricity competition are: (1) decreased confidence about investment efficiency under vertical integration, and (2) increased confidence in dispatch coordination among a large number of independent entities. The first factor is due both to the failures of the nuclear power industry and the emergence of an independent power industry based largely on highly efficient gas-fired combined cycle generation. The second factor is due to the enormous productivity gains in computation and control technology.

The crucial role of the transmission system, and the difficulties associated with providing transmission access remain the most important themes in the dialogue about the organization of electricity trade. These subjects have begun to attract the interest of economic theorists.

2.1.2 *Optimal Industry Structure*

Recent economic theory on the problem of optimal industry structure poses the issues as a regulatory choice. In this formulation, the regulator both chooses an industry structure and then enforces those rules or incentive schemes that are made necessary by the institutional framework. While this is a somewhat expansive view of the powers that any actual regulatory body may have, it allows a comprehensive treatment. In this section, we discuss briefly two recent studies that use this approach in order to characterize the problem formulation and to describe the kinds of results that are obtained. Vickers (1995) gives a general account of the trade-offs involved in regulating monopoly versus allowing competitive entry in vertically related markets. Gilbert and Riordan (1992) formulate the same general problem. To obtain tractable models, these authors must abstract significantly from the details of industry practice and technology. In the case where entry is allowed, both studies focus attention on the difficulties posed by the access problem. The vertically integrated firm must provide access to competitors. This is an inherently conflict-ridden role with the potential for inefficiency or abuse of market power.

Both studies adopt the modern approach to regulatory issues, which emphasizes the fundamental information asymmetry facing the regulator (Laffont and Tirole, 1993). Starting from this premise, the papers rediscover the basic trade-off between a vertically integrated structure and an industry structure which allows for competition in the segment where it is feasible. Vertical integration offers economies of scope across the upstream and downstream functions, but the information monopoly of the firm allows it to extract some rent from the regulator to achieve these economies. Allowing entry may sacrifice these vertical economies, but offers lower costs in the competitive segment. The loss of vertical economies shows up in the Vickers framework through excessive entry. Gilbert and Riordan emphasize the inefficient coordination across the upstream and downstream functions due to the inability of either the regulator or the disaggregated industry to operate the two functions efficiently on a separated basis.

2.1.3 *Estimating Vertical Economies*

A central question in the restructuring debate involves whether it will be necessary (or desirable) to divest the vertically integrated firm of its generation assets to achieve a workably competitive generation market, and what the costs of such divestiture might be. Given the basic trade-off between vertical economies and gains from competition, it is an important empirical question to develop some estimate of the magnitudes of these effects. Kaserman and Mayo (1991) address the issue of vertical economies. It is quite remarkable that these questions have been largely ignored in recent policy discussions.

Kaserman and Mayo (KM) formulate the problem by estimating cost functions for a sample of electric utilities, relying on accounting data, adjusted where possible to reflect economic costs rather than regulatory conventions. KM estimate a cost function for 1981 data, on a sample of 74 investor-owned utilities operating only in the electricity business. The argument for excluding firms operating in either the gas or water business along with electricity is that accounting allocations between these other businesses may introduce noise. The choice of 1981 data is based on the perception that earlier periods would be confounded by disequilibrium effects of rapid demand and input price changes. The sample includes 50 firms operating in both generation and distribution, 10 only in generation, and 14 only in distribution. The characterization of vertical economies involves testing whether firms that operate in both the generation and distribution segments have lower costs, all other things equal, than firms which operate in only one of these segments.

The statistical model that KM use (with several variations on the basic functional form) has the following general structure:

$$TC = a_1 + a_2 GEN + a_3 DIST + a_4 GEN^2 + a_5 DIST^2 + a_6 GEN \cdot DIST + b \cdot X \quad \textbf{(1)}$$

The dependent variable TC is the total dollar cost of electricity sales. GEN and DIST are the physical volumes of generation and distribution respectively. X represents a set of "control" variables, which are composed of various utility characteristics. These include fuel mix, labor costs, capital costs, customer mix, regional characteristics, and wholesale trading arrangements. The point of including these control variables is to "hold other things constant." There inclusion prevents the variables of interest, *GEN* and *DIST*, from picking up extraneous effects and allows us to state that, "other things held constant", *GEN* and *DIST* have a certain affect on *TC*.

The variable of central interest in this model is the "interaction" variable *GEN·DIST*. The coefficient on this variable answers the question: Does a firm with both generation and distribution produce more cheaply than a firm with just one or the other? If a firm has just one or the other, this variable is zero, while if it has both it is positive. Thus if $a_6 < 0$, the vertically integrated firm has been shown to have a lower cost than the separated firm.

The first four variables model the basic production process. The first two variables (GEN and DIST) model the fact that total cost increases roughly in proportion to the firm's size. The second two variables (GEN^2 and $DIST^2$) model the fact that marginal costs are increasing as output expands. In all of the models tested, the coefficients $a_2 - a_5$ are positive and significant. This indicates that marginal cost is positive and increasing, which is not unexpected, but which does contradict the assumption of a natural monopoly in generation.[1] But, like *X*, these variables only play a supporting role. They are included to pick up the basic properties of the production process just described and thereby keeping the interaction term free of contamination with these affects. Most importantly, the coefficient a_6 on the interaction term between generation and distribution is negative and significant at the 0.01 level in every model, which indicates the presence of cost complementarities between the vertical stages. This result is the principal basis for finding economies of vertical integration.[2]

Interpreting the statistical results, KM conclude that for a vertically integrated firm operating at the sample mean of generation and distribution (9,000 GWh and 7,200 GWh respectively), the costs of vertical disintegration are about 12% more than those of vertically integrated production. One key question is whether these results would stand up to further examination. As previously noted, the data are old (1981) and it is unclear if more recent data would change the results significantly. Perhaps more importantly, the KM study provides no clear indication regarding the source of these vertical economies. If these effects are large, it would be important to understand more completely where they originate.

The policy implications of these results are significant. The KM study strongly suggests that competitive restructuring may impose costs on electricity consumers. If the vertical economies are lost due to increased competition, then there should be efficiency gains of at least comparable magnitude in order for the change to be worthwhile. Alternatively, it may be desirable to retain the vertical structure, if possible, while enhancing competitive processes (see Chapter 6 for more detailed discussion).

---

1 The assumption of a natural monopoly for distribution does not rely on economies of scale as measured by the total amount of power sold, but rather as measured by the amount of electricity sold per geographic area. That hypothesis has not been tested. Note however that both of these results run counter to the notion that mergers are economic. These results indicate that the recent merger trend may have more to do with market power than with economies of scale.

2 It is also interesting that no other variables show up as statistically significant with any consistency across the various alternate model specifications.

2.2 Engineering Practice

In this section we describe the typical engineering practices that characterize the operation of the bulk power markets in the U.S. today. Historically, this activity has not been organized with commercial purposes, i.e., trade, as the predominant objective. Reliability issues, broadly construed, are the major consideration in engineering practice, particularly in the relations among utilities. Economic factors, however, are also significant determinants of engineering practice. Utilities seek to minimize costs within a framework of reliable operation. One way to think about the restructuring of bulk power markets is that it represents a shift in the balance between traditional reliability practices and economic objectives. Most of the reliability conventions adopted in the power industry arose out of informal practice, and not as an explicit optimization of reliability value.[3] To understand how these practices may be adjusted to allow for more commercial activity, we need to characterize current procedures.

2.2.1 *Reliability Councils*

Regional reliability councils were established in the U.S. following the Northeast blackout of 1965 to coordinate the reliability practices of utilities so that major disturbances could be avoided, or their impact minimized in the future. The reliability councils are voluntary organizations which engage in essentially two different kinds of activities. First, they establish minimum standards for operating procedures including issues such as the appropriate level of operating reserves, which includes both "spinning" and fast-start reserves. Reporting the performance of utilities with regard to meeting quality of service standards also comes under the area of operating standards. One important service standard is area control error (ACE). ACE measures the deviation between net power flow in or out of a utility control area and the net scheduled flow. The National Electric Reliability Council (NERC), the umbrella organization of all the regional reliability councils, requires that ACE be computed at least once every four seconds. Minimizing ACE is important for maintaining the standard frequency (60 cycles per second) of power supply. By encouraging utilities to meet the ACE standard, the reliability councils are limiting the extent to which one company is "leaning" on the grid to support network standards, rather than meeting these obligations themselves.

The reliability councils also engage in transmission planning activities, because these almost always have a regional aspect to them. The reliability councils have evolved into an arrangement of more or less permanent study groups that typically delegate to special subgroups the task of examining the regional and interregional implications of proposed new transmission projects (Fitzgerald and Hemphill 1993). An example of the results of these studies is the recently completed line ratings study for the new Mead-Adelanto-Phoenix 500

---

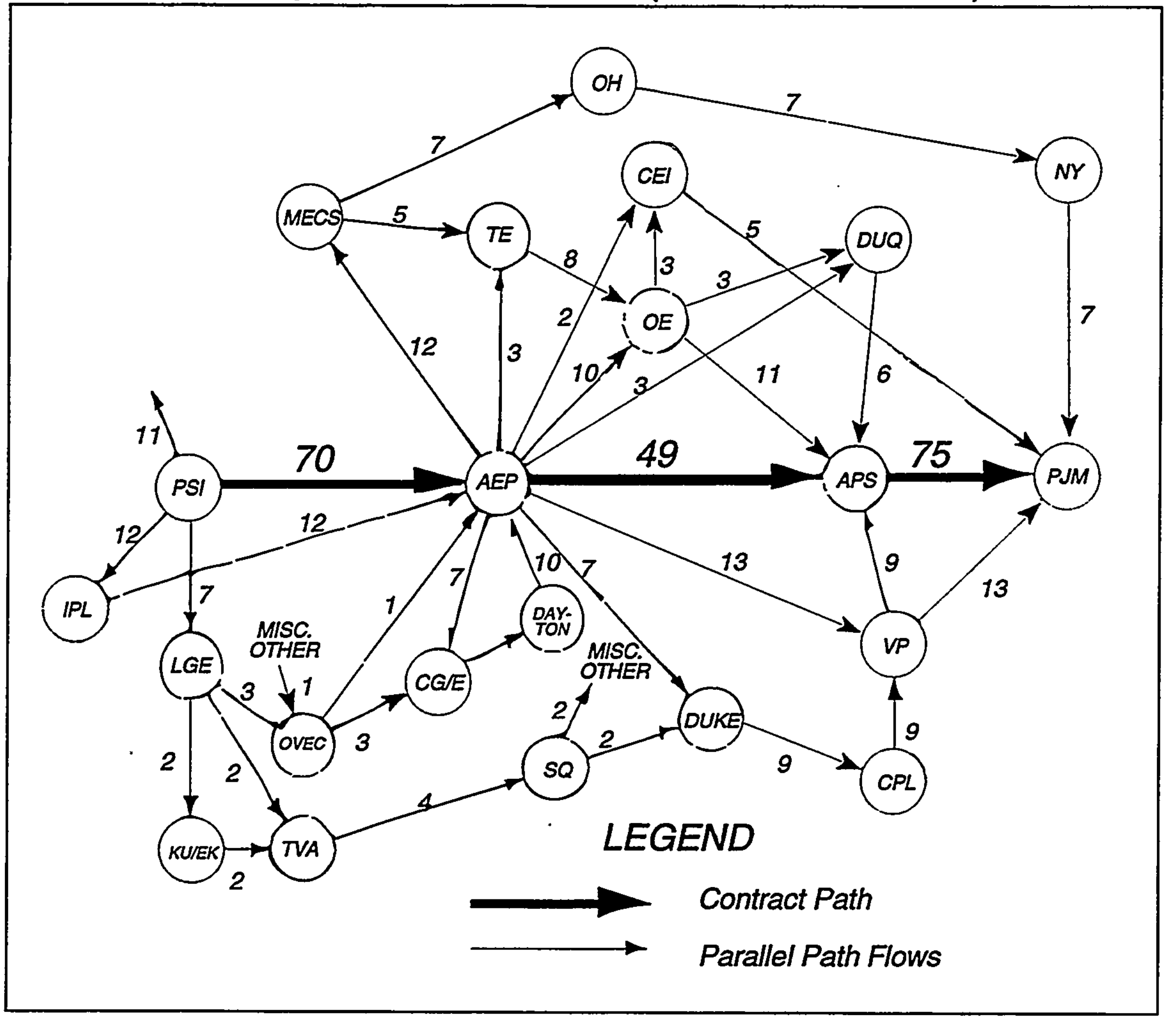
3 One example of the changing environment is the formulation of generation planning reserve targets on an economic basis, as in Southern Company Services (1991).

kV transmission project, scheduled to enter service in December, 1995. Line ratings are limits to the level of power flow that operators agree to maintain so that potential reliability impacts on the interconnected system are acceptable. Lee *et al.* (1995) describes the results of the study process leading to establishing the line rating for this project.

More recently, the reliability councils have gotten involved in pricing issues that involve network externalities. In electric transmission networks, actions in one location frequently have effects elsewhere. When the locations involved are under different ownership these interactions are economic externalities. Baldick and Kahn (1993b) give several simple examples of these in a planning context. In the operating context, the most common network externality is loop flow (sometimes called "parallel flow").

Figure 1, from Mistr (1992), illustrates the phenomenon. In this figure, a hypothetical 100 MW transaction is being simulated from PSI (now CINergy) to the eastern portion of the Pennsylvania New Jersey Maryland Interconnection (PJM). As this figure shows, the amount of power flowing over the assumed "contract path," from PSI to PJM, is only 50-75% of the 100 MW total. The rest flows on the systems of other utilities, not involved in the transaction. This loop flow phenomenon is common both in the eastern and western U.S. Utilities in both regions, through the reliability councils, and in special groups are attempting to develop compensation mechanisms to deal with this problem. Examples include the Western System Coordinating Council (WSCC 1994) coordinated phase shifter agreement filed with FERC and the General Agreement on Parallel Paths (GAPP) organized among utilities in the eastern interconnected grid. These examples represent the first time that explicitly economic issues have been addressed by the reliability councils.

Figure 1. Transfer Response 100 MW PSI to E PJM (1991 Summer Conditions)



2.2.2 *Operational Procedures at the Utility Level*

Electric utilities can be separated into those which physically control their own resources, and those which delegate that task to a larger regional neighbor. Of the nearly 3,000 electric utilities in the U.S., only about 160 operate their own control areas (FERC 1989; p.11).[4] This number of control areas is large by the standards of other countries; since it implies that the average amount of capacity controlled is about 4,000 MW. The control function can be separated into three broad categories: (1) unit commitment, (2) transactions scheduling, and (3) security constrained dispatch. This typology is not the only way to conceptualize the

---

4 There can be some ambiguity in the use of the term "control area." In some cases it may be used to denote less than the complete set of functions described in this section. We have made no independent assessment of the assertion cited regarding the number of control areas.

control function (see Kirby *et al.* 1995) for a more disaggregated view), but it will suffice for the purposes of this discussion. We briefly describe each of these activities.

Unit commitment refers to the operator's decision regarding the choice of generating units to turn on and off. This decision is usually made weekly, although in some cases it may be more frequent. Because there are significant thermal lags limiting the responsiveness of generating units, the start-up and shut-down schedule must be planned in advance. The factors affecting the determination of the optimal schedule include the forecast loads over the time horizon, the availability of units, and the operational constraints on those units. The unit constraints include minimum operating capacity, minimum up time and down time, ramping limits, and fuel limits. Using this information, the utility typically runs a computer program to decide what commitment schedule will minimize total costs. There is a great variety of algorithms available to solve this problem (Sheble and Fahd 1994). A good conceptual discussion with simple illustrative examples is given in Stoll (1989, pp. 410-419).

Figure 2 shows the unit commitment for Southern California Edison during a summer week in the mid-1980s. This figure shows only the utility's oil and gas units, which serve the intermediate and peaking segment, operating in addition to baseload coal and nuclear units.

The top line in Figure 2 represents the total nameplate capacity of oil and gas generation that is running and capable of serving load over this week. The line indicated by the open circles shows the actual generation from these units. Load fluctuates by approximately 6,000 MW from its highest to its lowest level during this week. The amount of capacity capable of operating (i.e., that is committed) fluctuates much less and much less frequently. During the weekdays (hours 1-108) the amount of capacity committed varies by about 1,500 MW. On the weekends, much more capacity is shut down.

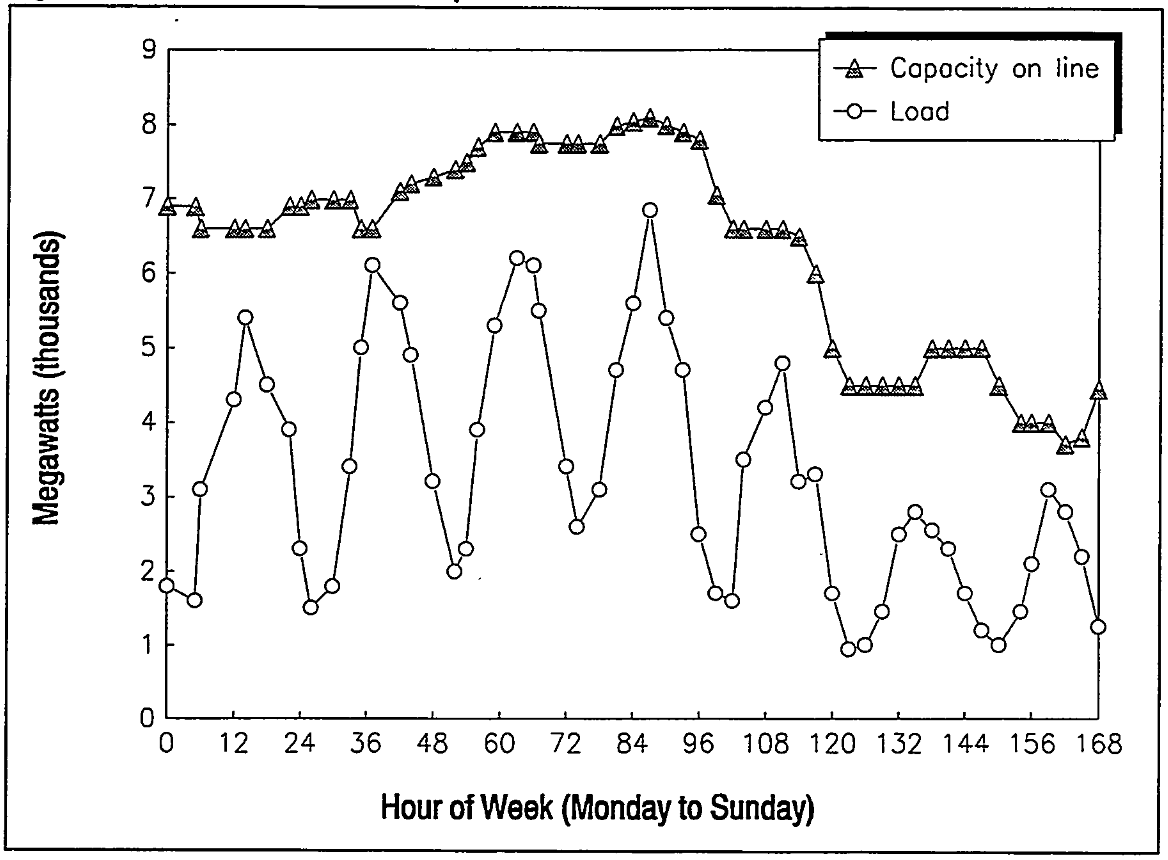
Figure 2 shows that substantial amounts of thermal capacity are committed but will operate at low output levels during low load periods. The efficiency of oil and gas-fired steam units at low output levels is poor. Nonetheless, it is more economic to incur this efficiency penalty than to start and stop units frequently.[5]

The unit commitment decision is made separately by each utility control center. One of the key questions involved in organizing bulk power markets to facilitate increased trade is whether there are coordination economies across control areas in the unit commitment function and if so, how they might be achieved. This question will be explored in more detail in Section 5.2.

---

5 Among the costs of frequent start-ups are increased maintenance costs due to stresses on equipment.

Figure 2. Unit Commitment and Dispatch



Transactions scheduling involves trade between utilities in different control areas. There are typically two kinds of transactions, firm and non-firm. A firm trade means that the supplier is responsible for meeting reliability requirements and cannot interrupt delivery unless his own system is in jeopardy. Firm transactions are accounted for in unit commitment decisions and security constrained dispatch. Non-firm transactions are scheduled over shorter time horizons and do not affect the unit commitment. Dispatchers may schedule non-firm transactions on a daily or hourly basis.

Security constrained dispatch (SCD) takes the available set of operating generation units as given and determines what the loading on each of them should be to meet the system loads. This decision is made within the bounds of reliability criteria established by the utility itself and the reliability council in which it operates. The reliability criteria include maintaining appropriate transmission line loadings, providing sufficient reactive power to meet voltage constraints, operating within stability requirements, etc. Once the reliability criteria are met, SCD chooses the generator loadings that minimize costs. In practice, the SCD is run at the utility control center every five minutes.

3 International Experience

Electricity restructuring has become a world wide phenomenon in the past decade. A broad range of new practices have been introduced in those countries that have reorganized the electricity industry. It is common to distinguish cases where liberalized entry has been allowed, typically in the form of independent power producers (IPPs) from more thorough reforms (Tenenbaum, Lock and Barker 1992; Smith and Klein 1994). Where institutional changes have involved reorganization of electricity trade, there may be some lessons for the U.S. We concentrate on these cases. In Section 3.1, we address the general issues affecting restructuring and identify those factors which distinguish the process in various countries. We emphasize the role of initial conditions. Section 3.2 reviews briefly the experience of individual countries and includes extensive references to the growing literature. In each country, the particular reforms strongly reflect the initial conditions. We do not draw out these dependencies in any detail, although common themes are identified.

3.1 Restructuring Processes

The restructuring process in electricity, as it has been implemented internationally, encompasses three general elements. First, vertically integrated firms are reorganized so that generation assets are separated from the transmission and distribution assets. We refer to this process as *functional restructuring*. It ultimately involves more than simply the divestiture of generation assets. Some kind of re-orientation of the transmission function is typically also required to facilitate increased competition among generating units and entities. Because the organization of the competitive process in generation can be so varied, it is convenient to refer to *generation competition* as a separate stage. Finally, electricity restructuring has frequently been accompanied by *privatization* of ownership. It is important to note that electricity industry reform originates from a situation of public ownership in most countries outside of the U.S.

Using this typology, Table 1 summarizes the restructuring process in a number of countries. The table highlights the fact that the entire process has been completed to varying degrees in various countries and that the U.S. industry starts from a substantially different point. Because of private ownership in the U.S., functional restructuring is linked to stranded cost recovery. Once the possibility of expanded competition arises, private investors need assurance of cost recovery for assets that would become economically obsolete if they are going to cooperate with such a transition. In the case of government ownership, losses are more easily absorbed. The cases of Spain and Alberta (Canada), where private ownership has a long history, involve recent functional restructuring (or serious discussion thereof) but limited competition to date. These cases are discussed in more detail in Section 6.2.

Table 1. How You Proceed Depends Upon Where You Start

Country	Functional Restructuring	Generation Competition	Privatization
UK, Chile, Argentina, Victoria (Australia)	⟶	⟶	⟶
Norway	⟶	⟶	
NSW (Australia) **New Zealand**	⟶	⟶ ?	
Ontario (Canada)	? ⟶		
U.S.		⟵	⟵
Spain, Alberta (Canada)	⟶		⟵

It is useful to recognize that different kinds of constraints may limit electricity restructuring. Jaccard (1994) provides an interesting discussion, in the context of varying conditions in the Canadian provinces, of the role played by technological endowments, particularly the dominant generation fuels. He distinguishes between "public goods" technologies, primarily nuclear and hydroelectric, and fossil fuel generation technologies, which are fundamentally private sector goods. Hydro resources clearly involve public lands and natural resource endowments for which there is a public interest. Balancing that interest with private concerns requires some government attention. Nuclear technology involves a number of externalities, both environmental and economic,[6] which are viewed positively in some countries (i.e., France) and negatively by at least some groups in other countries. Jaccard's point about the "public goods" technologies is that where they are predominant, there is likely to be public ownership, because private markets cannot easily internalize all of the associated externalities. Alternatively, reliance on public goods technologies may pose potential barriers to privatization.

6 The environmental externalities includes waste disposal problems, safety risks, and potential weapons proliferation. The economic externalities include the impact of developing a technological base that may create export opportunities.

Other important constraints on restructuring include the nature of local fuel markets, the strength of the transmission network, and the reliability performance of the industry. We discuss each of these issues briefly.

Competitive electric markets are facilitated by competitive fuel markets. To the extent that coal mines are vertically integrated with generation plants, the competitive situation may be complicated. Natural gas plays an important role in facilitating competition in many countries by lowering barriers to entry. Gas-fired combined cycle technology is relatively simple to build and operate and has low capital costs and very high thermal efficiency. Thus, where a well-developed natural gas infrastructure exists, competition will be more robust.

The transmission network is the vehicle through which electricity competition occurs. The stronger this network is, the smaller the limitations of network congestion. When the network is congested, competitive markets are geographically separated. This can be a problem for the efficient functioning of electricity competition. We discuss the role of transmission network constraints in Chapters 5-7.

Finally, reliability performance is an important dimension of the restructuring process. As a practical matter, no government would undertake electricity restructuring if it thought that the process would result in deteriorating reliability. In the developed countries surveyed below, it was always the case that the system was over built in generation. In the less developed countries, reliability performance was poor before restructuring. Improved productivity in these cases resulted in improved reliability. None of these situations, however, represent anything like a long run equilibrium. Whether competition produces adequate reliability in the long run remains unclear.

3.2 Individual Countries Experiences with Electricity Restructuring

The details of restructuring in different countries depends strongly on the initial endowment of generation resources, the political and economic culture, and the government policies that restructuring is designed to implement. The interplay of these factors is best understood through case studies (see Gilbert and Kahn, to appear).

3.2.1 *United Kingdom*

Newbery and Green (to appear) and Armstrong, Cowan and Vickers (1994) give comprehensive and balanced assessments of the restructuring history in the UK and its effects to date. The main limit on the extent of electricity privatization efforts in the UK was the nuclear assets. The financial markets would only accept these assets if the government was willing to provide broad guarantees that would have eliminated most risk and liability. The government refused to do this. As a result, nuclear generation remains in public ownership.

To implement generation competition, the UK has instituted a centralized spot market pool through which all electricity must be traded. There is a uniform price for all trades based on the bid price of the marginal unit. The level of capacity required is determined centrally by the pool's load forecast, with no demand-side bidding. The generation market is dominated by two large generating companies who have the ability to influence marginal price in the pool. Green and Newbery (1992) argued that the monopoly assets would have had to been divided up into at least five companies to produce effective competition. The large incumbent generators, National Power and Power Gen, dominate the market. They have engaged in a number of anticompetitive practices that have been documented in the reports of the Director General of the Office of Electricity Regulation (OFFER 1991, 1992a, 1992b, 1992d, 1993b, 1994). Among the more important of these practices has been manipulation of the pool price. The incumbent generators have also been able to manipulate transmission constraints in various ways, shifting costs onto all users through a pool surcharge called the uplift, which is added on to the pool price.

In addition to the spot market pool, there is a contract market operating between generators and users (large customers or distribution companies). This is strictly a financial market; we discuss its economic effects in some detail in Section 6.2.

Despite the lack of effective competition in the spot market, other aspects of the UK restructuring have had positive impacts. These include improved labor productivity, increased market share and profitability by Nuclear Electric, thereby improving the productivity of those assets, and the entry of new gas-fired combined cycle gas turbine (CCGT) technology (Yarrow 1994). There have been disputes regarding the extent to which this new entry was economic. Evaluating the need for new capacity in the UK is complicated because even though there was sufficient capacity even before the new CCGTs were added, impending environmental constraints would make much of it uneconomic (Newbery 1994). Most of the productivity benefits produced by the new system in the UK, however, have accrued to producers, rather than consumers; end-user prices have not declined.

3.2.2 *Norway*

The Norwegian system is 98% hydroelectric. The market was reorganized in 1991 by breaking up the vertical relationships, forming a separate transmission company, and instituting a trading regime. Electricity trade is dominated by physical bilateral contracts between buyer and seller that are dispatched by the central pool. This market is supplemented by a daily pool for all power that is not sold under contract. The pool price is set on a market clearing basis. In addition there are two other markets to meet technical and commercial requirements. The "regulation" market provides generation for the very short term balancing of supply and demand fluctuations. There is also a limited system of forward markets which can be used for price hedging (Knivsfla and Rud 1994).

The transmission network is sufficiently over-built, so that on average, no constraints limit trade (Hjalmarrson 1994). Nonetheless, transmission pricing accounts for potential congestion by charging a "bottleneck fee" based on the difference in spot prices on each side of a constrained interface (Moen 1995). This is the approach proposed in the spot market theory discussed in Section 5.1. The restructuring of the Norwegian industry involved separating the transmission network from the generating companies, but involved no privatization of generation assets; government ownership is still dominant. There had been a pool operation in Norway since the 1960s and a market-making operation since the early 1970s (Weideswang 1993), so the restructuring did not require creating completely new wholesale market institutions. The performance of the highly fractionated distribution segment of the industry is still believed to need improvement (Hjalmarrson 1994; Yarrow 1994).

3.2.3 *Chile*

The electricity system in Chile is a mix of hydro and thermal resources. The political and economic turmoil in Chile during the 1970s resulted in poor performance in the electricity sector and motivated the government to restructure. A systematic procedure of reorganization began with separating the state's commercial role from its policy-making and regulatory function. Next, a competitive wholesale market was established, and the vertically integrated monopolies were functionally separated. A regulatory system was established which included substantial pricing reform. Finally, the Chilean electricity industry was privatized. A substantial fraction of the assets are now under foreign ownership. The political background and details of the institutional development are described in Covarubbias and Maia (1994b).

The bulk power market operates through a centralized pool based on supply bids from generators that were spun off from the former monopolies and new entrants. Buyers are distribution companies and large customers. Bernstein (1988) describes the bulk power market institutions, including the role of the central dispatch pool and the pricing of transmission constraints. Galal (1994) gives a careful and positive assessment of the productivity performance of one generating company and one distributor following privatization. The Chilean system has the longest history of operation under a competitive regime, and was the conceptual model for the better known UK market.

3.2.4 *Argentina*

Covarubbias and Maia (1994a) characterize the productivity problems motivating the restructuring in Argentina, describe the institutional reform process, and provide some early results. The government proceeded with asset sales of both hydro and thermal generating stations in 1992 and 1993, but retained nuclear power under government ownership.

The generation mix in Argentina resembles Chile, except that the hydro resources are more decentralized and represent a larger share of total generation. The transmission network is more complex, and distances between generation and load centers are also greater than in Chile. Perez Arriaga (1994) describes the details of the new industry and regulatory structure. The dispatch operator who runs the pool depends upon decentralized hydro scheduling decisions made by project owners, whereas the pool controls the main hydro reservoir in Chile. Transmission capacity expansion decisions are also decentralized to some degree.

The pool structure in Argentina differs from Chile and the UK in two important regulatory dimensions (Perez Arriaga 1994). First, generators are not free to bid any price that they like. They are required to bid variable cost, which is subject to audit by the regulator. This feature may have been specified in light of concerns arising from experience with excessive bid prices in the UK pool. Second, the distribution companies purchase through the pool with a substantial lag. There is a "stabilization fund" which smooths out the price fluctuations that result from marginal cost pricing. This will limit the short term price responsiveness of distribution level demand. Large customers are directly transacting in the pool, however, and should be more price responsive.

3.2.5 *Australia*

The Australian power system is characterized by vertically integrated state-owned companies that serve each of the regional states. In the state of Victoria, whose capital is Melbourne, the industry has been vertically restructured and privatized along the lines of the UK. There is currently a wholesale pool operating in Victoria. New South Wales, whose capital is Sydney, is not as far along in the restructuring process. Some efforts have begun to increase the commercial orientation of the electricity generating company and the distributors, by removing them from direct government supervision and emphasizing profitability (GPT 1993). There may be a break up of Pacific Power, the generating company, in preparation for a more competitive market.

There have been experiments in interstate trade that were undertaken under the guidance of a federal agency (NGMC 1993a; NGMC 1993b). Due to the large distances between load centers and generating stations, the only practical trading regime would involve the states of Victoria, New South Wales, and South Australia. The trading rules currently under discussion will attempt to implement a decentralized market using the principles of spot pricing (Schweppe et al. 1988; also see Section 5.1), supplemented by forward market institutions (NGMC 1995). The purpose of developing forward markets is to improve the quality of the trading environment by helping to manage risk and improve the transparency of the market (Outhred and Kaye 1994). In Section 5.5 we discuss the current NGMC market rules in the context of unit commitment coordination economies.

The significance of the NGMC initiatives for the U.S. lies in their "bottom up" nature. The federal authority over the companies in each state is limited, so that industry consensus is required for significant change to occur. The decentralization of the industry parallels the current organization of the electricity industry in the U.S. This institutional setting contrasts strongly with the situation in most other countries, where "top down" restructuring has been the rule.

3.3 Summary

In Chile, Argentina, and the UK, a centralized pool operates the wholesale market, with all trade going through the pool. A bilateral regime dominates the Norwegian market, but it depends upon some unusual conditions. First, there are very limited transmission constraints in the network, so that many possible dispatch arrangements are feasible. Secondly, before restructuring, there had already been a framework of pooling and marketing institutions, which facilitated a smooth transition to a competitive wholesale structure. The UK, Chile and Argentina all formed their pooling institutions from the single area control function performed by a central utility operator. The Australian NGMC experiment promises greater market decentralization than the other cases. This will be facilitated by having relatively few market centers involved in the trading regime.

Our analysis suggests that, overall, the electricity restructuring processes in other countries offer few major lessons for the U.S. because of pervasive private ownership and the highly fractionated nature of the U.S. electricity industry. Private ownership is important because it makes the process of reorganization of functions more complex compared to situations that involve government enterprises exclusively. If restructuring requires writing down the value of uneconomic assets, this is more easily accomplished by governments than if private investors must be compensated. The large number of relatively small utilities in the U.S. power network means that the formation of pools would also involve complicated financial compensation procedures.

For the current institutional framework of the electricity industry in the U.S., there may be insights from the Australian NGMC experiments where decentralized trade coordination will be tested. The other country experiments offer potential lessons in pool economies, operation and transaction costs. Both Argentina and Chile are trying to decentralize transmission system expansion, which should also provide valuable lessons.

4 Current Structure of the Bulk Power Market

The economics literature reviewed in Section 2.1, insofar as it addressed empirical issues, is significantly out of date with regard to important features of current bulk power markets in the U.S. In this chapter, we focus on issues that affect the prospects for efficient trade. In particular, we address the role of non-utility generation (NUGs), market price formation and the performance of existing bulk power market institutions.

NUG issues are addressed in Section 4.1. The emergence of non-utility generators reflects the success of PURPA's policies to facilitate entry. However, in many areas of the U.S., entry by non-utility generators has also reduced operational flexibility for utility system operators, which results in increased operating costs. Understanding the current role of most non-utility generation in the existing bulk power market is important because it highlights the point that ease of entry for new participants does not necessarily or inevitably lead to a more efficient electricity industry overall.

In Section 4.2, we discuss the importance of price transparency in the development of competitive markets. We then examine obstacles to price transparency, specifically some of the practices of existing U.S. institutions, such as power pools, which are responsible for coordinating and facilitating trade in bulk power markets. We also highlight the efforts of various market participants to develop and produce wholesale market price indices in response to the increasing demand for price transparency.

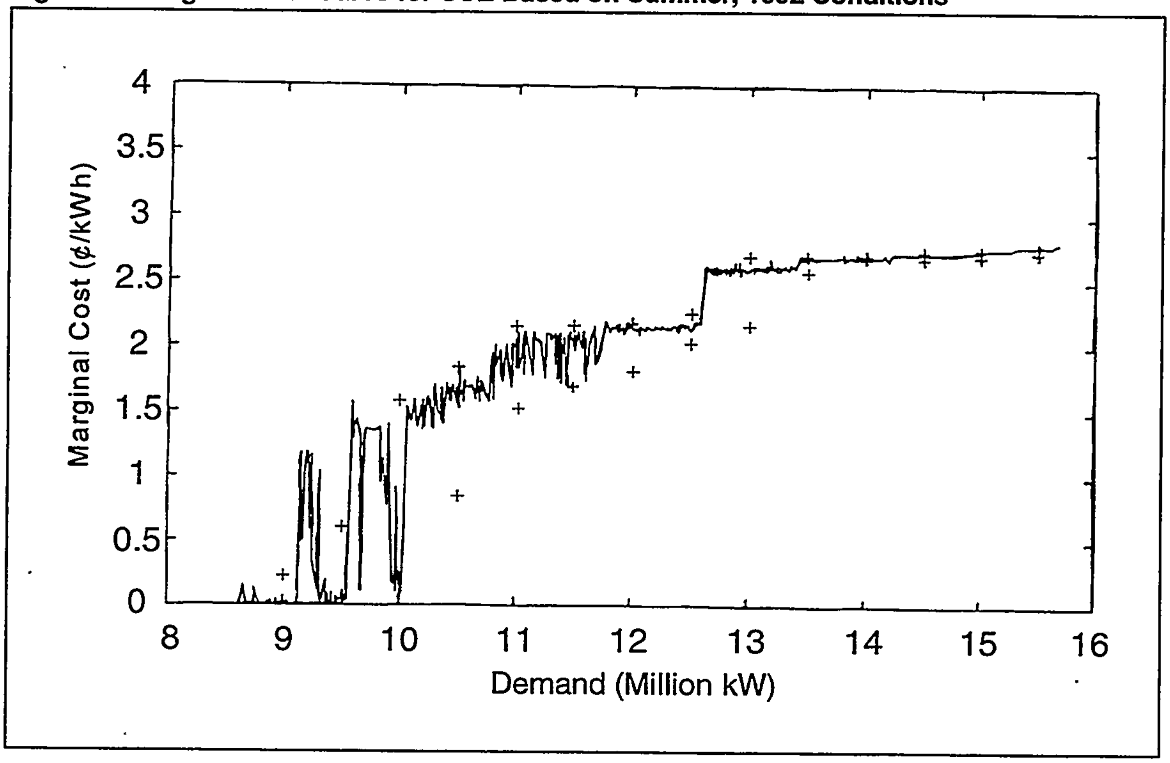
In Section 4.3, we survey the performance of U.S. bulk power market institutions, focusing on power pools. Our review indicates that power pools have produced demonstrable benefits for utilities by reducing excess capacity, thus leading to reduced reserve margins. However, the studies also suggest that there are not significant efficiencies in dispatch for utilities in power pools vs. companies that are not in pools and engage in bilateral trade. Based on current practices in the U.S., there is not definitive evidence regarding the superiority of voluntary bilateral trade vs. pooling arrangements.

4.1 Non-Utility Generation

There is a fundamental asymmetry between utility responsibilities and NUG access rights in the bulk power market. Because PURPA mandates that utilities must take power from Qualifying Facilities (QFs) under almost all circumstances, the burden of making operational adjustments to accommodate changes in the supply/demand balance falls entirely on utility-owned generators. In regions such as California and the Northeast, these are important factors. Ilic *et al.* (1993) describe procedures used to manage random fluctuations in supply and demand in this setting. The burden of cycling generator output under expected conditions is equally, if not more important. As the output from a generator varies, and as the number of start-ups increases, there is additional wear and tear that requires O&M expenses. Due to the growing perception of these issues, NUGs acquired competitively in recent years are obliged to provide operating flexibility and are being evaluated along that dimension (Kahn, Marnay and Berman 1992). NUG contracts in Virginia, for example, exhibit substantial operating flexibility, which has resulted in much more adaptation to unexpected conditions than is typical elsewhere. But the competitive segment is still small compared to the much larger population of non-dispatchable NUGs. Data on the recent performance of dispatchable projects and the relative size of this segment is given in Comnes, Belden and Kahn (1995).

There are no general estimates of the costs associated with the asymmetry in operational obligations, but there is reason to believe that these costs are significant. Some of the costs are monetary. These are primarily opportunity costs resulting from must-run PURPA production. Utilities must purchase high priced QF output when cheaper alternatives are available. Other costs involve reduced reliability, such as a lessened ability to meet Area Control Error standards. Figure 3, from White (1994), shows the minimum load problems of Southern California Edison in the form of extreme marginal cost instability during periods of low demand. The price spikes at the low end of the cost curve show cases where generators start up and operate at their minimum stable levels, but these levels are sufficiently large that they may exceed demand, or at least reduce opportunity cost to extremely low levels. It is quite likely that these cost instabilities at low demand result in lower performance

Figure 3. Marginal Cost Curve for SCE Based on Summer, 1992 Conditions



on Area Control Error, but NERC does not make data on this question publicly available. Other examples where supply inflexibilities result in operating problems associated with generator cycling are discussed qualitatively in Le *et al.* (1990).

One approach to quantifying the cost impacts of QF must-run operation more generally would be to examine regions, such as California, New England or New York, where the NUG impact is large and examine what the savings in operating costs would be if the QFs were dispatched. The difficulty with such an approach would be getting reliable estimates of the true QF variable costs. Alternatively, a lower bound would involve comparing the outage rates of utility generators with those of NUGs, as well as the change in utility outage rates as NUG penetration has increased. The implicit argument of such comparisons is that the adjustment costs of increased cycling requirements show up as reduced reliability of utility generators. It would be much more difficult to estimate the costs of the control area reliability case discussed in connection with Figure 3.

Some reform of the NUG market is currently being debated by federal and state policymakers. Changes in this regime need to simultaneously respect NUG contract rights, but also encourage re-negotiation to increase operating flexibility.

4.2 Price Transparency

Transparency of various kinds has become an increasingly important regulatory objective as the process of infrastructure deregulation has developed. Transparency is one aspect of the larger question of information management and disclosure which is a subtle problem in regulation. The fundamental proposition of the modern theory of regulation is that the monopoly firm has an important information advantage over the regulator (Laffont and Tirole 1993). Competition tends to erode the information advantage of the monopolist. This occurs both by the functional restructuring process, which requires information about cost, and by competitive processes, which tend to produce information about price. At the same time, however, that deregulation increases pressure for information transparency, there are pressures for information protection. Competitive advantage for individual firms is frequently involved. If a competitor must disclose all relevant market information, then there are no returns to developing it.

As a working definition of the transparency concept, we use the simple criterion that something is transparent if it can be known to any interested party at nominal cost. We can speak about transparency of prices, of costs and of processes. Price transparency may be the simplest of these notions. The reason that price transparency is desirable is to facilitate competition by making apparent the transaction price for various products. In countries where such prices are either secret, or at least difficult to obtain, regulatory reform may begin with requirements for price transparency. This appears to be the case in much of Europe (De Paoli and Finon 1993).

Transparency of process is more complex, if for no other reason than that processes are complex, whereas prices are inherently more simple. Decision-making is frequently obscure in state-owned infrastructure industries. Often, no one has to explain why a particular decision was made and therefore no one can be held accountable. This is in marked contrast to systems (even government owned) in which a regulatory agency must give an account of its decisions. Increasing the transparency of government processes is a common objective of deregulation (Smith and Klein 1994). Firms frequently employ complex processes to optimize their own operations in either monopolistic or competitive markets. The need for transparency here is much less, since firms are accountable. Only in cases where complex processes affect the competitive position of other participants can a case be made for the public value of process transparency in the individual firm.

Finally, cost transparency has several dimensions. In a competitive market, prices approximate costs closely. In the case of regulated firms with dominant market positions the situation is more opaque. Regulators can observe costs through auditing. This may be expensive, but it is feasible. Where costs are jointly incurred to produce more than one service or product, the separation is basically arbitrary. In this case cost observation is essentially infeasible. The marginal costs of some products produced by dominant firms are difficult to observe because they result from complex (i.e. non-transparent) processes.

Price transparency is an important element of the competitive process because it reveals the value of products that might not otherwise be known either by market participants or by regulators trying to facilitate the transition to a competitive market structure.

In the following sections, we survey various current practices that are obstacles to price transparency as well as procedures that may facilitate it. First, we examine power pool operations organized along central dispatch lines through a case study of the New York Power Pool (NYPP). The translation of the engineering rules used to operate central pools into commercial terms can be difficult. Next, we examine the recent development of market indices, which characterize transaction prices at various locations in the U.S. power network.

4.2.1 *Power Pool Practices and Price Transparency: NYPP Case Study*

In the U.S., there are three "tight" power pools, which are located in the northeast: the New England Power Pool (NEPOOL), the Pennsylvania-New Jersey-Maryland (PJM) Pool and the New York Power Pool (NYPP). Utilities in other regions of the U.S. have entered into "looser" pooling arrangements, none of which involve central dispatch. To illustrate current practices of power pools as they relate to price transparency issues, we focus on the New York Power Pool, because its operation has been particularly well-documented in several studies (Ruganis 1986).

The NYPP, like other centrally dispatched power pools in the U.S., operates on a "split-savings" pricing rule. This means that the marginal costs are not charged to buyers and paid to sellers, but rather that an intermediate price is used for transactions which reflects what the buyer would have paid. This pricing policy introduces an extra level of complexity into conventional U.S. pool pricing and contributes to making the pricing more opaque. The opacity stems both from the specific nature of the price for every buyer for every transaction and from the lack of a mechanism to make this information available broadly. Needless to say, it would be quite difficult to audit or verify. In contrast, pools implemented in other countries rely on marginal cost pricing (see Section 3.2).

Ruganis (1986) is a very careful and interesting, if somewhat dated, study of the operating cost structure of the New York Power Pool (NYPP). Ruganis has available to him the operating records of the NYPP, and seeks to construct a characterization of the hourly marginal costs. While his purpose is to develop data that can be used for benchmarking a multi-area production simulation model, his study has implications beyond that purpose. Among other lessons, this study illustrates the difficulties of using engineering data to characterize the commercial aspects of electricity trade. In formal pools, such as the NYPP, the transactions price between parties has frequently been based on the "split savings" concept. This requires a separate calculation of the hypothetical cost of buyers, which has many of the same conceptual problems as marginal costs, but does not have the property of being observable in principle. Thus, the very fact that retrospective marginal cost studies are necessary indicates the basic point that centralized dispatch may not produce readily transparent transaction prices. We summarize this study to illustrate the gap between power system operations and electricity trade.

System operations in the NYPP are governed by a security-constrained dispatch (SCD) model that allocates minute by minute fluctuations in load to the generators on the system which are operating. Ruganis characterizes the principal objective of SCD models as maintaining system stability; economic optimization is secondary to that primary objective. SCD models yield shadow prices that can be interpreted as marginal costs. Ruganis shows, however, that these values are too finely grained to be usable for estimates of the "market price" of electricity. That is, the SCD shadow prices can fluctuate substantially over short time intervals, and their values are quite sensitive to the length of the time interval over which they are measured. Further, SCD shadow prices only measure very small marginal changes, and so can neglect the impact of units that provided "most" of the adjustment to a load change observed over a somewhat longer time horizon. As a result there may be "noise" in SCD estimates of marginal costs. It is not clear if this implies a biased estimate. It is clear, however, that the microstructure of marginal cost is less significant for commercial purposes than the larger view based on discrete transactions.

The NYPP uses available hydro resources along with its predominantly thermal generators. The scheduling and valuation of storage hydro resources presents another major problem in the translation of engineering to commercial practice. The economic principle on which the valuation of storage hydro resources is typically done is by estimating the opportunity costs associated with its use. This means that its value is the cost of the resources that it would have displaced at some future date, had in not been used today. The problem is quite complicated when formulated over a multi-year horizon. The complexity involves deciding how to allocate water storage and use over these time cycles.[7] Hjalmarrson (1995), for example, describes this problem in the Norwegian setting. The competing generators, all of whose resources are hydro based, use stochastic dynamic programming models of the power system to design bidding strategies in light of hydrologic conditions, a demand forecast, and their assessment of the behavior of competitors. For such analysis to be even feasible, there must be common models of the hydrology, the power system, and demand characteristics. This is certainly not a transparent process, but there is no reason for it to be. All that matters is that the pool price which emerges from competition is transparent to users. There is neither a transparent marginal cost nor a transparent transaction price in the U.S. "split savings" power pools.

Even in systems where hydro plays a much smaller role, the basic problem remains that over any time horizon longer than a month, the estimation of storage hydro value is quite uncertain. Therefore any hydro scheduling plan is quite likely to be suboptimal after the fact, and any forecast of marginal costs is highly dependent on the storage hydro schedule. This means that marginal cost depends upon the administrative decisions of the hydro scheduling program, and this is a non-transparent process.

4.2.2 *Wholesale Market Indices*

Although there is a very large volume of wholesale electricity trade, the transaction prices are not easily available. Firm transactions have typically been regulated on a cost of service basis by FERC. They are reported in the annual filings of investor-owned utilities to FERC. Short term non-firm transactions are also reported on an annual basis, but at a level of aggregation that does not reveal the structure of transaction prices. The FERC reports of non-firm trade simply add up all the transactions between two parties that occur over a year and report them as one sale or purchase.

Recently, market participants and private information service providers have begun to cooperate in efforts to produce more transparent indicators of wholesale transaction prices. All of these efforts are designed to produce price indices that reflect market activity in different products at particularly active trading locations in the electricity network. One of these indices is being developed by Dow Jones/Telerate reflecting activity at the California-Oregon border (COB). The COB market index will cover four different electricity products:

---

7 The problem resembles the operator's decision to invoke the limited curtailment rights obtained from customers analyzed by Oren and Smith (1992). In both cases, a fixed quantity of a valuable resource must be allocated over a future period where the occurrence of the periods of maximum value is uncertain.

firm and non-firm energy delivered during peak and off-peak hours (Speckman and Schleimer 1995). Other providers of market price information include newsletters such as *Power Markets Week* and *California Energy Markets.*

It is not clear exactly how these indices are computed or what their accuracy may be. We can illustrate the index construction issues with a simple example. Table 2 gives four hypothetical trades characterized by a price and a quantity. The two obvious ways to construct a price index is to compute a simple average price or a quantity-weighted average. For this example, the simple average is 31 and the quantity weighted average is 31.8. In this case, therefore the choice of averaging method turns out to make only a small difference. More generally, this would not be true. Clearly, the quantity-weighting method is conceptually preferable.

Table 2. Hypothetical Trade Data

Trade	Quantity	Price
A	5	20
B	10	40
C	20	35
D	30	29

The potentially larger problem with indices is sampling bias. The data used in the index may represent a biased sample of the underlying market because the data collection is not based on a random process. Some market participants may not provide their data, thus introducing a potential distortion in the reported index price. Using Table 2 data, suppose the transaction sample only included A and D. Then the simple average price would be 24.5 and the weighted average would be 27.7. In this case, sampling error introduces a large bias, although the error is less with the quantity-weighted index.

Another important issue involves the question of what product is being described. The current indices appear to be focused on very short-term products, typically one day ahead schedules. However useful such products may be, they may not be the main products transacted. Electricity production planning over longer time cycles is much more important than daily planning. Unit commitment schedules typically involve a one week plan. Fuel purchasing typically involves a minimum of one month planning. Therefore, electricity price indices for transactions representing these longer time horizons may be more economically meaningful than the day ahead products.

Price indices are useful measures of economic value. At their current state of development, it is likely that the indices have sampling bias problems. As the wholesale market develops, these problems will probably be diminished. Competing information providers will have an incentive to demonstrate the quality of their products by trying to assure more complete sampling. It remains to be seen whether price indices develop for electricity products traded over longer time horizons than the day ahead schedule.

4.3 Performance of Market Institutions

Empirical studies of market performance have seldom yielded conclusive results. A subject of traditional interest is the relative performance of public vs. private ownership. Pollitt (1994) is a recent study of this kind. We focus on the performance of trading institutions. Studies of these institutions have been typically conducted at a fairly high level of aggregation, because it is necessary to observe many trades to arrive at any conclusion about the functioning of commercial practices. Qualitative studies also play an important role in isolating particular practices that may limit performance. Such studies also can be fairly far removed from engineering practices.

The performance of U.S. power pools has been examined recently from a number of different perspectives. Pechman (1993) and the staff of the New York Public Service Commission (NYPSC 1991) examine the functioning of the New York Power Pool from organizational and information perspectives. Both identify serious impediments to improved efficiency in the NYPP due to the unanimity rule governing changes in procedures. One area where efficiency could be improved involves the operation of the single pumped storage unit in the state. Each member utility has certain scheduling rights for this unit. The result of decentralized decision-making is less optimal use of the facility as a whole compared to pool-wide operation. More beneficial arbitrage is limited in favor of schedules that improve the position of only one company. Another example of a coordination economy that the NYPP does not capture is single area unit commitment. Unlike other centrally dispatched pools, the NYPP does not coordinate the commitment of units among the utilities.

Gilbert, Kahn and White (1993) examine data on power pools to determine if there are observable efficiencies associated with pooling. They examine the level of operation of baseload units owned by utilities that participate in pools compared to those that do not. In principle, pooling should increase the output of low cost baseload units by "flattening" the load curve that these units serve. This means that there should be a better utilization of their capacity than for non-pooled units. Data for 1989 suggests that this is, in fact, the case. Closer inspection, however, suggests that the observed differences are due primarily to differences in installed capacity. Utilities that do not pool typically had greater excess capacity at this time than those that pooled. Once the operational data is adjusted for the differences in capacity, there is no longer any observable difference in performance. These results, while they fail to support the claims of operational superiority of pooling, do support the other main coordination claim of pooling, namely that it results in better planning.

Finally, White (1995) studies the potential for increased pooling in California. He takes a detailed look at the benefits of dispatching hydro resources, located primarily in the North, across the entire state. His simulation concentrates only on the peak month, so that operational benefits such as joint unit commitment or coordinated maintenance scheduling are excluded. He also ignores transmission costs and constraints. The result is an estimated benefit of about 4% of total costs, due principally to lower costs of peak load production. The net benefit, once transmission effects are taken into account would be less. These results are consistent with the belief that wholesale markets in the Western U.S., operating under the currently accepted set of constraints, are not grossly inefficient. It remains to be seen, however, whether larger efficiency gains are possible if some of the current constraints and trading practices were changed.

The implications of these results for competitive markets are not entirely clear. If competition results in the reduction or elimination of barriers to trade, then operational efficiency should increase. It would appear that the opportunities for increased trading efficiency would lie in the interaction of non-pooled utilities either with one another or with existing pools. In the short run, competition will probably induce increased trade based on under-utilized assets. In the longer run, competitive pressure may or may not facilitate efficient investment. The pools seem to have achieved less excess capacity than non-pooled utilities. It is reasonable to expect that competition will reduce excess capacity in the intermediate term. Long-run investment in a competitive market may not be adequate (see Section 6.1.3 for discussion of this issue).

5 Trading in a Competitive Power Market

As activity in the wholesale market has increased, and state regulatory commissions have begun seriosly discussing deregulation of the bulk power market, a variety of views have been offered on the best way to structure this market. Most of these views propose a fully competitive structure, but these views divide into two quite distinct categories. The nodal pricing models would require all trades to be made with a central market maker, the ISO. In contrast, the bilateral models specify that trades should be made between private parties. Both sides recognize that an ISO is necessary to control the externalities associated with trading over a common network, and both sides claim that their proposals will best approximate the workings of a competitive market.

In this chapter, we do not attempt to resolve this debate, which is occurring in many regions and countries and will undoubtedly persist for years. Instead we describe the basic nodal pricing model and three variations of the bilateral approach, and explain some of the central arguments for and against the two generic approaches.

Section 5.1 presents the influential nodal spot market theory stripped of its mathematical complexity. Section 5.2 shows how nodal spot prices provide a basis for several long-term contracts that can be used to eliminate some of the risks associated with spot markets.[8] Section 5.3 considers a criticism of the nodal approach that claims trading only with the ISO is too restrictive, while Section 5.4 asks whether we need a mandatory pool at all, or if the market can self-organize efficiently without the ISO being directly involved in economic dispatch. This section also summarizes three academic versions of the "bilateral" approach. Section 5.5 discusses forward contracting in a bilateral regime. Finally, Section 5.6 examines the coordination economies of a centrally dispatched pool.

5.1 Spot Market Theory

The technological characteristics of electricity networks have led to the development of a pricing and trade theory based on a particular concept of spot markets. Originally articulated by Schweppe (1978), and fully developed in Schweppe *et al.* (1988), this approach has been revived more recently by a number of authors in their proposals to restructure bulk power markets. Accounting for network properties is a fundamental requirement of any electricity trading regime. The spot market theory gives particularly strong emphasis to network properties. Because this theory has been thoroughly studied, it can be more clearly defined than the more recent bilateral proposals.

---

8 In Section 6.1.2 we examine whether the full structure of both spot and contract markets might mitigate the market power of dominant participants.

5.1.1 *Spot Market Definition*

Spot market theory defines the optimal spot price at each node in a very simple and natural way, although one often finds that definition replaced by one based on complex procedures for calculating that price. We will not concern ourselves with the details of calculation, and will only give the underlying economic definition.

Definition: The optimal spot price (per kWh) at a node is the minimum system cost of supplying one more kWh at that node when the system is optimally dispatched.

The optimal nodal spot price is a very useful concept because it measures the true cost of supply and as a consequence provides the conceptually correct price to charge a demander. If demanders are presented with this price, they will consume up to the point where their marginal value of power is equal to the spot price. Thus the optimal nodal price will also equal the value of supplying a marginal kWh at the node in question.

In defining optimal nodal spot price, we also introduced the concept of "optimal dispatch," which we now define.

Definition: The optimal dispatch maximizes system net benefit, i.e. the difference between the total benefit to customers and the total cost of generation.

Among power system planners, this is often referred to as the "optimal power flow," or OPF. However in calculating an OPF, planners typically assume that demand is fixed, which simplifies the demand side of the problem. OPF is easy to define but hard to calculate. In fact one of the central controversies surrounding nodal spot pricing proposals is the difficulty of computing the OPF. Recent experience reported by PG&E highlights this issue (Papalexopoulos *et al.* 1994). OPFs are generally very sensitive to both the precise specification of system constraints and to slight differences in the marginal cost of generation. Thus, there are typically a set of power flows that dispatch very different generation sets than the OPF, but that are nonetheless extremely close to optimal in terms of system net benefit. This makes it possible for the system operator to favor one generator over another without being readily detected. For this reason, independence of the system operator is desirable. This property of OPFs has been cited by Wu and Varaiya (1995) as a critique of the spot pricing model. We discuss their arguments in Section 5.3.2. The importance of this issue has not been settled conclusively.

5.1.2 *Properties of the Spot Market*

To gain a deeper intuitive understanding of optimal nodal spot prices, we now discuss several of their properties. Spot prices are partly determined by the supply and demand for power, but they are also affected by two properties of the transmission grid itself: losses and constraints. Our first property describes how prices would be set if these two properties were not constraining. In other words we assume a grid without losses or congestion.

Property 1: In a loss less and uncongested grid, the optimal nodal spot price will be the same at all nodes, and will equal every generators short-run marginal cost and every demanders short-run marginal benefit.

In this case the entire system forms a single perfectly competitive spot market. While there are many times when a grid is uncongested, there are always losses, so we consider that case next.

In the spot-price regime, the independent system operator (ISO) is the buyer to every seller and the seller to every buyer: all trades are made with the ISO. When considering this aspect of the ISO's role, we generally refer to the ISO as a grid merchant, and refer to the grid merchant's net gains from trade as the "merchandizing surplus." Property 2 considers the part of the merchandizing surplus that arises from the spot price differences that correspond to losses.

Property 2: In an uncongested but lossy grid, the optimal nodal spot price is lower at generation nodes and higher at demand nodes. These differences are great enough to earn the grid merchant a merchandizing surplus, which is approximately equal to the value of lost power.

This property is somewhat surprising. One might have expected that if 10% of the power is lost when shipping from node A to B, then the price at B would be 10% higher. This would imply that the system operator would just break even. However, power losses are proportional to the square of the power shipped, so the marginal loss is twice the average loss. Consequently, the price at B, which reflected the marginal loss, will be 20% higher than the price at A, and the grid merchant will make a profit of 10%.

We now turn to the even more subtle effect of congestion on spot prices. To simplify this discussion we again assume a loss less network. This assumption is not problematic, because the nodal spot price differences caused by losses and those caused by congestion are simply additive in the complete model.

Property 3: In a loss less grid with one congested line, the optimal spot price will be higher at the end receiving power by an amount that measures the value of increasing the line's capacity.

If the line from node A to B is congested, then it is not possible to increase consumption at B simply by increasing generation at A, for this would cause the power flow on line A-B to exceed its rating. For that reason, the cost of supplying power to B will be greater than the cost of generating it at A; thus the difference in nodal spot prices.

Beyond this point the effects of congestion become quite complex. To give the reader some idea of the nature of these complexities we state two more properties that demonstrate the counter-intuitive nature of real power flows.

Property 4: With only one congested line, it is possible that for some other line in the network the spot price at the demanding end will be lower than at the generating end.

This appears counter-intuitive because we generally expect that electric power will only be bought at node A and sold at node B if it is cheaper at A than at B. Because of the apparent backwardness of this transaction the corresponding power flow is often called an "uphill flow."[9] This possibility has been most closely analyzed by Wu *et al.* (1994).

Possibly even more disconcerting than the standard uphill flow described in property 4 and by Wu *et al.* is the possibility of an uphill flow on a congested line, which we describe in our final property.

Property 5: In a grid with more than one congested line, the optimal nodal price at the receiving end of a congested line may be higher, lower, or equal to the optimal nodal price at the transmitting end.

In spite of these complications, the optimal spot price differences between nodes do send the correct economic signals to both generators and demanders. Both are properly discouraged from contributing to system losses and from contributing to congestion. Thus, if they can be computed by the system operator and utilized by traders, they will play exactly the role one would hope for them to play; counter-intuitive properties not withstanding. These commendable properties are simply the result of optimal prices being defined as the shadow

---

9 Figure 7b in Section 6.3 shows an example of an "uphill" power flow. In this example, line 1-2 is congested and is preventing node 1 from transmitting additional power to node 3. But if, as happens in Figure 7c, node 2 transmits additional power to node 3, this causes a counter flow on the congested line and makes it possible for node 1 to inject more power. If we assume the prices at node 1 and node 3 are 5¢ and 10¢ respectively, we can find the optimal nodal spot price at node 2 as follows.

The optimal price at node 2 is the net system benefit of an additional kWh. Every kWh injected at node 2 has a direct benefit of 10¢ to node 3. But every kWh injected at node 2 also has an indirect net benefit of 5¢ because it allows node 1 to ship an additional kWh to node 3. Thus, the value and price at node 2 is 15¢/kWh, and the flow from node 2 to node 3 is an "uphill" flow: it takes power from an expensive node to a cheap node.

prices of generation at each node. In other words, optimal spot prices, by design, take into account all system costs and benefits.

5.1.3 *Spot Market Implementation*

This conceptual discussion of spot market theory is sufficiently abstract that it requires further institutional specification before a trade regime can be defined coherently. The spot market model can be implemented to varying degrees of approximation. One particularly important question involves the issue of system dispatch. Who will operate the network to produce the flows that link the prices at different nodes? The early formulations of this question were somewhat vague. For example, Hogan (1992) assumes that there need be no change in operations from existing institutional arrangements to implement spot market prices for use as transmission prices:

> "...transmission prices can be estimated ex post....The ex post method allows the current dispatch operations to remain in place and calculates prices consistent with the actual usage by applying the marginal tests of economic dispatch." (p. 224).

This approach requires "accepting the actual system dispatch as an optimal balance of the underlying economics and constraints..." (p. 223). For a regime of limited trade among vertically integrated utilities, this approximation *might* be acceptable. In a regime of more expanded trade where marketers and perhaps end-use customers are transacting, this approximation will not suffice, because the dispatcher has incentives to favor trades that benefit his own units and not those of competitors. In any particular situation, many dispatch patterns are feasible, although they affect access opportunities differentially. Therefore, the dispatcher can discriminate in response to network constraints.

More recent discussion of spot market theory has situated it institutionally in a centralized pool setting (Garber, Hogan, and Ruff 1994). In this implementation, concerns about biased dispatch are mitigated to a considerable degree, although not eliminated (see Wu and Varaiya 1994).

5.2 Spot and Contract Markets

The spot market provides for the real-time matching of supply and demand, which is a unique requirement of the electricity industry. However, spot markets alone are not likely to meet other requirements of an efficient market. By its nature, spot prices generate a risky income stream, which can be problematic as a basis for long-term investment. By design, the spot market can provide opportunities for manipulation of the process, and therefore of the resulting spot price or prices. These problems (i.e., risky income stream and opportunities to manipulate prices) can create barriers to entry, which in the long run undermine the competitiveness of the market.[10] These problems can be mitigated by long-term financial contracts, which are the subject of this section.

To motivate our discussion on the use and value of long-term contracts, we develop a stylized example that involves a pair of traders who wish to make a long-term trade at a fixed price. This example allows us to introduce three basic types of long-term contracts which can be overlaid on a spot market: a contract for differences (CFD), a transmission congestion contract (TCC), and a forward contract (FC).

The first and simplest step for the supplier and demander is to write a CFD. This can be done without any support from a long-term contract market; it is entirely private in nature. A CFD eliminates the temporal risks associated with the spot market, but not spatial risks, as it does not account for the price difference between nodes. A second step is needed to completely eliminate risk from the long term contract, and this step is taken by purchasing a TCC. This requires a market in TCCs, for there is no way for the traders to originate such a contract on their own. Having both a CFD and TCC completely eliminates long-term risk. As Table 3 illustrates, part of the TCC cancels, in a mathematical sense, part of the CFD, and the result is equivalent to a forward contract (FC). This allows us to conclude that a forward contract could be used in place of both CFDs and TCCs.

In Table 3, we illustrate these relationships more concretely with a supplier at node I and a demander at node j that decide to trade a quantity of power q, at the fixed contract price P_c. In a true spot-market regime this trade could not be made directly, but instead each must trade with the ISO; the payment from this transaction is shown in row 1 of the table.

10 When there is a competitive fringe interacting with vertically integrated dispatcher, there are incentives to find an equilibrium that damages competitors (Kahn, 1995a). As the experience in the UK amply illustrates, an independent system operator is no guarantee of efficient dispatch.

Table 3. Using Long-Term Contracts to Eliminate Nominal Price Risk

	Contract or Market	Payment	
		Supplier at Node *I*	Demander at Node *j*
1	Spot Market	$P_i \cdot q$	$-P_j \cdot q$
2	CFD for q at P_c	$(P_c - P_j) \cdot q$	$-(P_c - P_j) \cdot q$
3	Total	$P_c \cdot q - (P_i - P_j) \cdot q$	$-P_c \cdot q$
4	TCC for q from I to j	$(P_i - P_j) \cdot q$	---
5	Total	$P_c \cdot q$	$-P_c \cdot q$
6	FC sold at I, bought at j	$-P_i \cdot q$	$P_j \cdot q$

The easiest step in moving from the spot market to a long-term fixed price contract is to implement a contract for differences, as described by row 2. This contract can reference either node's spot price or even the average of the two. We have chosen one typical and convenient implementation for our example, and have defined the CFD to require the demander to pay the supplier the difference between the contract price, P_c, and the spot price at node *j*. As can be seen from row 3, if the two spot prices are the same, this contract exactly transforms a series of spot transactions at uncertain future prices into a single long-term transaction at a fixed price, thereby eliminating all price risk of nominal price fluctuations for both parties. If the parties are equally risk averse, this elimination of risk is accomplished without the payment of a risk premium by either party, something that cannot be accomplished by a forward market involving outside speculators. If one party is more risk averse than the other, that party should still pay less for risk reduction than in a futures market because the other party can provide that reduction at no cost.

Because a nodal spot price market does have nodal spot price differences whenever intervening lines are congested, the complete elimination of price uncertainty requires the use of another type of long-term contract. For this purpose Hogan (1992) invented the transmission congestion contract (TCC).[11] As shown in row 4, a TCC pays its bearer the spot price difference between the specified nodes times a fixed contract quantity. Hogan has proposed that the investors who build the grid should be rewarded with a TCC, which is issued and backed by the ISO. This again has the advantage of eliminating risk without the need to pay any party a risk premium, but it would still be possible for an outside financial institution to make a market in TCCs.

11 Nodal prices also differ because of losses. The standard definition of a TCC ignored this complication and we will continue in that tradition, but the interested reader who consults Bushnell and Stoft (1995b) will find that TCCs can be extended to lossy networks with only minor modifications. The principle one being that they must be broken into two parts and consequently come to look exactly like a pair of forward contracts.

As can be seen from row 5, which sums rows 3 and 4, the addition of the TCC completes the process of transforming a series of spot trades into a fixed price long-term contract. This is seen by the fact that the payment shown in row 5 is fixed; it does not depend on any spot price. The last line in the table shows the payments from forward contracts at the two nodes: a short contract at *I*, and a long contract at *j*. These contracts exactly cancel the spot-market transactions shown in row 1. Thus, if the demander buys a forward contract at *j* and the supplier sells a forward contract at *I*, they have accomplished the same thing as is accomplished with a TCC and CFD. This point is made by Oren *et al.* (1995).

This concludes our brief overview of the mechanics of various types of long-term contracts; the role of long-term contracts in addressing the shortcomings of spot markets is discused in more detail in Section 6.1.2.

5.3 Will a Mandatory Pool Restrict Bilateral Contracts?

A concern of those who favor "physical" bilateral contracts is that the obligation to trade with the Pool prohibits the implementation of useful "physical" bilateral contracts, thereby reducing market efficiency. PoolCo advocates point out that those engaged in a bilateral contracts can easily insure that their generator will be dispatched and their load will be fully served simply by manipulating their bid prices. Because a contract for differences (CFD) insulates them from spot price fluctuations, and because they will trade at the spot price and not at their bid prices, the outcome of their trade with PoolCo is exactly the same both physically and financially as if PoolCo had been absent from the transaction.

Bilateral advocates claim that there are exceptions to this equivalence. One example that has been pointed to in discussions is the case of a contract between a hydro generator and a load that wants to contract for a fixed amount of energy without specifying the delivery time. Although it is admitted that a CFD would be equivalent to a physical bilateral contract in the absence of congestion, it is claimed that congestion will drive a wedge between the two outcomes. Although it is true that CFDs do not eliminate the risk of locational variations in the spot price, it is also true that parties with a physical bilateral contract should have to pay these same congestion charges. The claim that trading with a PoolCo is more restrictive than bilateral trading certainly deserves more attention, although so far the argument has not been fully articulated and supporting evidence is limited.

5.4 Do We Need A Mandatory Pool to Compete?

Various types of bilateral trading models represent the major alternative to the nodal spot market implementation of a competitive bulk power market, sometimes referred to simply as the Poolco approach. These bilateral trading models can be distinguished from PoolCo proposals in three ways:

> (1) The ISO does not take title to the power during a trade; i.e. the ISO is not the buyer to every seller and the seller to every buyer.
>
> (2) The ISO does not have control over the dispatch except to the extent needed to maintain system integrity.
>
> (3) There are no nodal prices; spot prices are always prices for bilateral trades, which may sometimes be spot (short-term) prices, but which are not associated with a single node.

In California, several bilateral trading models have been put forward by various parties in response to the California Public Utilities Commission's (CPUC) "Blue Book" proposal. We will not try to recapitulate the specific proposals put forth in California, many of which have evolved through the workshop and political negotiation process. For discussion purposes, we focus on three academic models that span a large part of the spectrum of "bilateralist" positions. The three academic models are relatively simple to summarize and are internally consistent, if perhaps less practical, than the ones put forward in the regulatory arena.

The three models are Wu and Varaiya's "coordinated multilateral trading" (CMT) approach (1995), Chao and Peck's transmission-bidding approach, and McGuire's transmission-charge approach. All three models are quite recent, appearing for the first time in 1995; each model has some areas that have still not been well-defined. It also must be noted that Chao and Peck's model is not simply bilateral but envisions the possibility that bilateral trades taking place along side spot-price trades with a PoolCo. Nonetheless, it does show one way a competitive bilateral/multilateral market could be organized, and we will focus only on the bilateral trading mechanism.

5.4.1 *Chao and Peck*

Both the CMT approach (Wu and Variaya), and the transmission bidding approach (Chao and Peck), remove the ISO not only from the energy market, but also from the market for transmission services. This does not mean that the ISO has no roles in these markets, only that the ISO does not buy or sell transmission services. In both approaches, the ISO gives information to the traders on the extent of losses that are attributable to their transactions on the network. This includes not only the power lost in their own shipments, but also any losses that they impose on other trades. The traders are then responsible for covering the losses they cause, either through generation of their own, or through purchases. Of course, it may happen that a trader fails in this responsibility, but at that point, we are discussing imbalances rather than losses. Imbalances are handled by the ISO by procedures that are not intrinsic to the systems under discussion. Other ancillary services are also handled by the ISO based on pre-specified procedures.

The most interesting and definitive aspects of these two models lie in their treatment of congestion. In the transmission-bidding approach, transmission rights to lines are owned privately, and must be acquired before a trading party can use the line. The ISO has a major role to play in this regard. Since the path of power flow through a network is determined by the laws of physics and not by the traders, or even by the ISO (barring the use of control devices such as phase shifters), a given trade will typically flow over a large number of lines; in fact it is not technically incorrect to say that to some extent it flows on every line in the entire system. The ISO computes these flows for the trading parties and the trading parties are then required to secure the appropriate level of rights on all the affected lines. As Chao and Peck readily admit, this is a daunting task, at best requiring state of the art telecommunications and computing equipment.

It is also important to note that the market for transmission rights *must* include a rule that forces the ownership of each line to be spread among many parties. If not, and the rights to even one line were owned by a single party, that party could stop all trade on the network by refusing to sell that right to the traders that need it. Of course such stark monopoly power would earn the owner a handsome income. From this example we see that moving to a bilateral trading system does not necessarily remove all problems of market power.

5.4.2 *Wu and Varaiya*

In contrast to Chao and Peck, Wu and Varaiya have gone to great length with their CMT approach to reduce the computational burden needed to "coordinate" their multilateral trading market. Their ISO provides traders with very simple formulas that embody the system losses and system security limits with respect to any set of traders. Traders are required to cover the losses they cause and are not required to obtain rights to use network lines. Traders are simply required to respect the system limits.

In situations in which a number of non-cooperating traders each want to use a single line and their total usage would exceed the lines' security limit (based on the system contingency analysis), the ISO must then resolve this conflict. Wu and Varaiya acknowledge that the ISO's method for resolving this conflict is "arbitrary."[12] By this, they appear to mean only

---

[12] Private correspondence from Felix Wu to Steven Stoft (7/22/95).

that their lemmas and theorems apply irrespective of the curtailment rule used by the ISO. However the curtailment rule does matter to the traders.

To understand the potential impact on traders, we must consider the concept of optimal congestion. In each model (i.e., bilateral, multilateral, and PoolCo), when the system works as planned by the authors, there are market forces that induce traders to optimally adjust their trades to the congestion of the system. This means that in all of these systems, each trader must experience, through some market mechanism, the negative external impact on other traders caused by his own trades' contribution to system congestion. How does this happen in the CMT approach? When trades are initially proposed for any half hour time slot, the ISO curtails any group of trades that exceed system limits. This means that the ISO arbitrarily divides up the rights to the congested lines and gives them to the traders who submitted bids for those lines. Being scarce, these rights are valuable, and may well be sold to parties who can make better use of them. If they are sold, that imposes the correct congestion cost on the buyer. If they are not sold, that imposes the correct congestion cost on the owner in the form of an opportunity cost.

It is interesting to note that in a lossless network, the value of the curtailment rights that are "arbitrarily" handed out by the ISO are exactly equal to the "merchandising surplus" collected by the ISO in a PoolCo regime. This is a consequence of the fact that the Wu-Varaiya system constrains the merchandising surplus to zero, yet to, achieve efficiency, must in effect distribute these rights to the traders. If these rights are curtailed in a way that reflects energy use or sales, then we believe that curtailment will introduce a distortionary incentive for either use or generation, thereby causing inefficiency. Wu and Varaiya deny this claim.

5.4.3 *McGuire's Transmission Charge Approach*

We review briefly a bilateral proposal that has been developed by McGuire (1995), which is less well-known, but is formulated in very practical and down-to-earth terms. McGuire (1995) asks the question "Is power really so special?" and answers it with a firm no. As a consequence, McGuire thinks it should be traded like any other commodity: two traders make a deal, then they go to the shipping company, pay for shipping, and complete their transaction. He admits that it is more difficult to compute the cost of shipping electricity than other goods, but still finds this only a small problem. He also admits that the shipping (transmission) company has a natural monopoly, and so he proposes a way for them to compute transmission charges that will make it easy to regulate them. Having done that, he proposes to leave the business of trading to the traders, as we do in every other market.

McGuire's system is simple in its structure. Traders submit proposed trades, i.e. 10 MW from node I to node j, to the ISO. The ISO provides information on curtailments and transmission prices. Trades are now allowed to take place. But if traders want to, they can revise their trades based on the transmission prices they have received, and the ISO will reschedule and return new transmission prices. If the traders persist in resubmitting, then McGuire's rules for the ISO will lead to an equilibrium in which no trader finds his trade curtailed by the ISO, and no trader wants to propose a different trade. The challenge in this system is in specifying the rule by which the ISO adjusts the price of transmission. Since this rule is specified by the regulator, and the ISO has no choice as to how it is applied, the ISO cannot impose unfair transmission charges. The primary difficulty of the system is technical. For each one hour dispatch, a set of iterations between the traders and the ISO is required. At this time a good estimate of the number of iterations required is not available. The costs of the iteration process may be prohibitive.

5.4.4 *Comparison of Bilateral Trading Models*

Our provisional assessments of these three schemes may be summarized as follows. The Chao-Peck scheme requires such a complex auction that it is probably far beyond the realm of workability. Their approach also presents serious market power problems with respect to transmission rights that have not yet been addressed theoretically and may well be unsolvable. Wu and Varaiya's scheme appears workable, and it may do more to help traders arrive at an efficient set of trades than McGuire's does, but it also seems to have potential problems with the allocation of transmission rights. McGuire's system seem superior to Wu and Varaiya's because it does not allocate transmission rights arbitrarily, but instead sets a price for it and allows any that are willing to pay that price to have access. This would seem to prevent the gaming and arbitrariness that would be an inevitable part of the Wu-Varaiya system.

Table 4. Comparison of Three Bilateral Trading Models

<table>
<tr><th></th><th>Chao & Peck</th><th>Wu & Variaya</th><th>McGuire</th></tr>
<tr><td rowspan="2">Role of ISO in Trades</td><td rowspan="2">Report actual power flows on all lines as result of proposed trade.</td><td colspan="2">Arbitrarily curtails trades in excess of rated line capacities.</td></tr>
<tr><td>Provides information about loss and congestion.</td><td>Imposes transmission charges on traders.</td></tr>
<tr><td>Energy Trade Prices</td><td colspan="3">Strictly confidential between trading partners</td></tr>
<tr><td>Energy-Trade Quantities</td><td colspan="3">Reported to ISO</td></tr>
<tr><td>Transmission Rights</td><td>Privately owned. Auctioned off every period.</td><td>Allocated arbitrarily by ISO in each trading period.</td><td>Owned by ISO who charges traders for losses & congestion.</td></tr>
<tr><td>Losses</td><td colspan="3">Computed by ISO, covered by traders.</td></tr>
<tr><td>Imbalances</td><td colspan="3">Handled by ISO.</td></tr>
</table>

5.5 Forward Markets for Bilateral Traders

As discussed in Section 5.2, forward markets can be useful for hedging the risks of nodal spot prices, but their role in bilateral markets has somewhat different dimensions. The most fundamental difference is that the bilateral models assume that forward contracts close with physical delivery of the product. They are not only financial instruments, but entail what contract lawyers call "specific performance." Because of the specific performance feature, it has been argued that a PoolCo is not such a good idea because it would be hard to establish a forward market in such a setting. These ideas have been articulated most completely by Levin (1995a, 1995b).

Levin brings a unique perspective to the discussion of electricity trade because he represents the New York Mercantile Exchange (NYMEX) which is the leading futures exchange for energy products. NYMEX developed a highly successful futures market for natural gas which matured along with the deregulation of that industry. NYMEX has recently submitted proposed electricity futures contracts for approval by the Commodity Futures Trading Commission. The issues associated with an electricity futures market are addressed in Belden and Kahn (1995). Here we concentrate on Levin's views of the economic role played by forward markets.

Levin observes that forward markets are in fact the dominant form of exchange for most commerce. Buyers and sellers benefit from planning their business activities in advance, and therefore predominantly contract for future delivery. The role of spot markets is primarily to allocate supply in light of planning errors and random shocks on the supply or demand side. This is necessarily a smaller segment of total market activity than that which is represented by forward markets. As a result, Levin argues that the prices in forward markets will typically be lower than those in spot markets for the same good. This basic relationship is the reason that Levin believes that using spot market pools as the required medium of all electricity trade is inefficient.

In a spot market pool, where all electricity trade must clear, forward markets are reduced to financial instruments. While pool proponents assert that this will provide adequate risk hedging opportunities for participants who do not wish to rely on spot prices exclusively, Levin argues the contrary position. The key difference that Levin perceives between forward market trading with physical delivery responsibility and the strictly financial forward contract is that the former will be lower cost than the latter. A seller in a mandatory pool environment would have no incentive to contract forward at a price that is less than the expected spot price. A seller's incentive in a market structure without such a pool is to gain customers. This incentive is lacking where all demand clears through the pool.

In addition to his analysis of forward contract inefficiencies in a mandatory pool, Levin also provides an interesting analysis of potential regulatory distortions that might arise in a mandatory pool. Since a mandatory pool will have enormous influence over the structure of the electricity market, there will be an interest in regulating the behavior of participants. For example, in Argentina, bidders are required to bid their true variable cost to the pool, not just a price that they choose strategically; they are subject to audit on this (Perez Arriaga 1994). The goal of this requirements is to avoid some of the distortions in the UK pool. The result of such a requirement, however, is potentially to limit the fuel pricing flexibility of sellers. Levin (1995a) lists six different fuel pricing conventions that might be mandated in a regulated "merit order dispatch" regime.[13] In a decentralized trading regime, individual agents would decide on whatever fuel pricing they deemed was appropriate and take the corresponding risks.

Levin's discussion abstracts completely from the technological constraints in electricity markets, which are the foundation of spot market theory. His positive view of forward market contracting is consistent with the kind of decentralized trading discussed in Section 5.3, but does not really address how the appropriate level of multilateral coordination might be achieved. Finally, he does not take the possibility of transmission manipulation to be a serious issue; neither this nor any other form of market power enters his discussion. In Chapter 6, we examine a few of the ways in which electric networks give rise to opportunities for exercising market power.

5.6 Coordination Economies

Spot markets may be implemented through pooling institutions or in a more decentralized fashion. The central dispatch argument for pooling emphasizes very short run coordination economies. With improved communication and control technologies, it is unclear whether centrally dispatched pools will achieve any significant improvement over decentralized trade. A more significant potential source of coordination economy involves intermediate term reserve sharing, specifically unit commitment and maintenance scheduling. Short term trading among wholesale market participants is based on hourly or perhaps day ahead costs. The intermediate term scheduling of maintenance, unit commitment and reserves operates over time horizons of one week to one year. Here the coordination problem is more difficult because it may be difficult to get decentralized agents to contract with a high level of certainty to achieve the benefits, or to devise compensation schemes that will share the benefits.

The magnitude of coordination economies in the intermediate term is difficult to estimate. One recent study which addresses unit commitment benefits is Lee and Feng (1992). We discuss this study in some detail. Lee and Feng consider three utilities connected by transmission links of varying capacity. The greater the transmission capacity connecting them,

---

13 These include: (1) "swing" gas delivered the next day, (2)"swing" gas delivered the next week, (3) "spot" gas delivered the next month, (4) the average price of gas imputed from a twelve month commodity swap, (5) the opportunity cost of re-selling gas in the cash market, and (6) any of various definitions of historic purchase costs.

the more they can share capacity (no attention is paid to network effects). There are two impacts of increased coordination. First, fewer units (or more efficient units) need to be committed. The benefit of that is reduced inefficient operation at minimum load. With fewer units operating, however, there are fewer opportunities to trade among the utilities. We report the net effect of these two different impacts in Table 5 and disaggregate the net impact in Figure 4.

Table 5. Multi Area Unit Commitment Economies

Transmission Capacity		Case A Coordinated Unit Commitment ($000)	Case B Coordinated Dispatch Only ($000)	B-A	
MW	% Area Peak			($000)	(B-A)/B
0	0	3002	3002	0	0
50	3.8	2918	2960	42	0.014
100	7.7	2818	2938	120	0.04
150	11.5	2761	2927	166	0.056
200	15.4	2691	2924	233	0.079
250	19.2	2649	2924	275	0.094
300	23.1	2626	2924	298	0.101
350	26.9	2615	2924	309	0.105
400	30.8	2612	2924	312	0.106
450	34.6	2611	2924	313	0.107

Figure 4. Commitment Economies vs. Lost Trade Opportunities

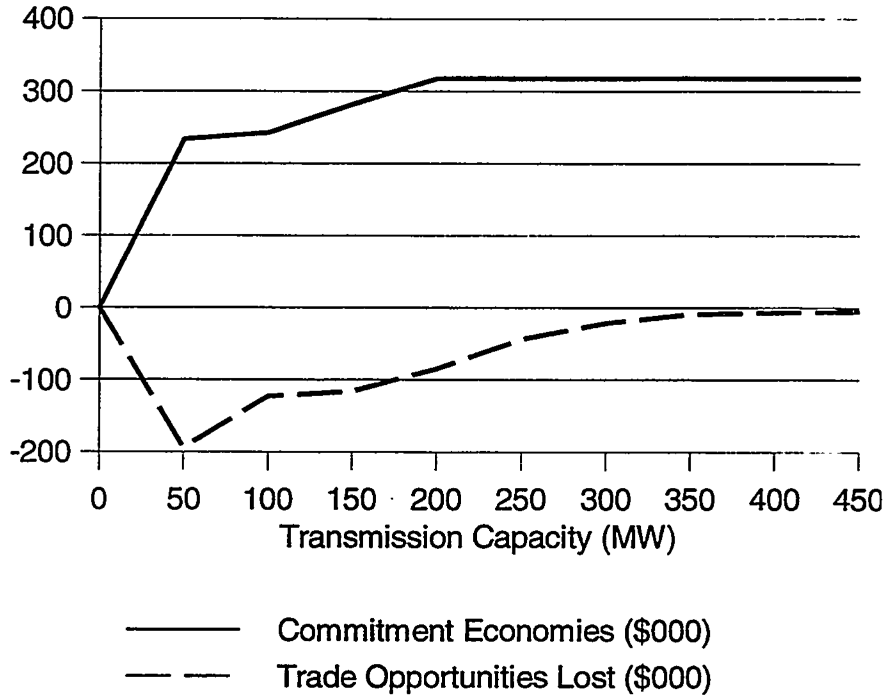


Table 5 summarizes the simulation results as presented by the authors. Case A represents the situation where unit commitment is coordinated among the three areas and they are jointly dispatched. In Case B, each area commits units separately and then there is joint dispatch. The sensitivity of results to the magnitude of transmission interconnections is illustrated in the table. Transmission capacity is listed in MW interconnections between any two of the utilities. To give a better idea of total transmission requirements for these cases, we also express the total transmission capacity as a fraction of the total peak demand for all three utilities.

Presented in this fashion, however, the underlying trade-offs are not apparent. To illustrate those, we disaggregate the simulation results into the benefits of joint commitment, which is reduced inefficient operation at minimum load, and the reduced trade opportunities that result when fewer units are committed. To make this separation we first identified the units from the EIA *Inventory of Power Plants* and estimated their unit price for fuel from the EIA *Cost and Quality of Fuels* report. Data on the unit input/output functions and minimum generation levels is provided in the study. Figure 4 summarizes the trade-off between unit commitment economies and diminished trade. This figure shows that the avoided cost benefits of coordinated commitment saturate more quickly than the trading opportunities are achieved.

The reason that multi-area commitment is beneficial is that the minimum operating levels of the units are highly inefficient. The average heat rate for this data is about 12,000 Btu/kWh for the baseload units, while their full load heat rates range between 9,500-10,000 Btu/kWh.

This study does not address the optimal level of transmission capacity, but from the results in Table 5, it appears that marginal benefits beyond the level of 250 MW interconnection among all participants are rather small.

It is an open question whether pooling institutions will be required to achieve coordination economies of this kind, or whether a more decentralized market structure can achieve the same result. The NGMC experiment in Australia is the most serious effort internationally to implement decentralized trading. The NGMC Code of Conduct (1995) specifies that it will be the responsibility of individual generators to decide when and whether to commit their units. These decisions are subject to advance notice requirements and various measures of system reliability. A short term forward market will be set up with the intention of providing generators with an opportunity to minimize the risks associated with a commitment decision by locking up a price for at least their minimum output, or perhaps for more than that. This forward market is intended to be a purely financial market.

6 The Problems of Market Power

When competition is imperfect, firms may be able to exercise market power, thereby distorting trade and potentially reducing efficiency. The sources of market power vary from industry to industry. In this chapter, we examine key factors and/or situations that may create market power in electricity: market concentration, vertical integration, and the ability to block transmission pathways.

Market concentration is the typical source of market power and will probably prove to be the most decisive factor in bulk power markets as well. In Section 6.1, we address the question of concentration and explore several related concepts. We show that forward contract markets tend to ameliorate the effects of market concentration, while congestion increases the level of market concentration. Our analysis also suggests that a limited amount of market power derived from concentration is found to play a helpful role in maintaining reliability. Perhaps the central question concerning the organization of U.S. bulk power markets is whether vertical re-organization of the entire industry will be required. Section 6.2 addresses the problem of market power engendered by collusion between the DistCo and GenCo halves of vertically integrated firms. Specifically, we discuss the potential for this collusion to inhibit entry by other GenCos and the possibility that an ISO can control this type of behavior. In Section 6.3, we examine the possibilities for Gencos to block transmission access for strategic purposes. In Section 6.4, we discuss and critique recent transmission access policies put forth by FERC in its Mega-NOPR. We argue that FERC's open access objectives simply cannot be met successfully without the use of an ISO, and that even with one, the pre-determined transmission prices envisioned by FERC impede the implementation of an efficient access rule.

6.1 Market Concentration in Generation

In this section, we review briefly techniques used by economists to analyze market concentration as applied to bulk power markets. We then focus on two areas that are, if not completely unique to electricity, at least quite unusual and generally not well understood: the effect of congestion on market size and the effect of competition on reliability.

6.1.1 *Standard Issues in Market Concentration*

Considering generation first, the central question is whether one firm will have a large enough share of the market to profitably raise the market price significantly. This question can be approached via the standard Cournot oligopoly analysis, which tells us that the answer will depend on the elasticity of demand, and the market share of the generating firms. Specifically, the markup of a firm with share-of-market s and elasticity ϵ, will be given by:

$$\frac{p - c}{p} = \frac{s}{\epsilon}. \qquad (2)$$

On the left we have the fraction by which price exceeds marginal cost. As an example, assume that a certain firm has 20% of the market, and the long-run elasticity of demand is 0.8, which is quite elastic. In that case we would expect a markup of 0.2/0.8 = 0.25, which means that 25% of customer payments go to covering markup, while 75% cover marginal costs. This is quite a substantial markup. More recently, contestability theory has been used to argue that the firm will price lower than this in order to prevent the entry of competitors. An even more pertinent correction to the Cournot model has been pointed to by a number of authors who argue that bilateral contracts will dramatically reduce markups in the spot market. This is discussed at some length in the following subsection.

Market concentration may also be a problem in the distribution market, the demand side of the market. We do not analyze buyer market power, which is referred to as "oligopsony power," in detail, but offer several reasons why we think it is not likely to be too important. Most importantly, although a Disco may have a large market share it cannot make a centralized decision to curtail demand the way a large generating company can decide to curtail generation. This is because the Disco typically serves many customers who would find it impossible to collude on their consumption decisions. A second limit on oligopsony power is the elasticity of the supply curve. Just as the market power of an oligopolist is limited by the elasticity of the demand curve, so an oligopsonist's market power is limited by the elasticity of the supply curve. Because entry costs in generation are low and because gas-combined cycle technology exhibits relatively few economies of scale, the long-run supply curve appears to be quite flat, i.e. extremely elastic. Because of this, reducing demand won't lower price very much.

6.1.2 *The Usefulness of Forward Contracts*

Forward contracts replace the uncertainty of the spot market with the certainty of a fixed price contract, which is highly desirable for an investor in the long-lived physical assets of a power plant. Forward contracts help assure investors that a generation project will be able to sell its product at a reasonable price over a sufficient time frame for investors to recoup their investment. From a financier's viewpoint this makes the project safe; from the investor's viewpoint, this makes his project "bankable." Thus forward contracts facilitate entry into the generation market. In the long-run, this will limit the market power of the current players, and in the short run it will dissuade them from exercising some of the market power they have, because they do not want to encourage entry. There are other ways in which forward contracts can mitigate market power which also deserve attention.

A number of writers have identified a second potential benefit of forward markets. This is as an important mitigant of market power in the spot markets to which the forward markets are linked (Allaz and Vila 1993; Green 1993; Powell 1993; Newbery 1995). In the context of market power, an important conclusion of these papers is that forward contracts reduce the exercise of market power in the spot market. After presenting the basic argument for this position, we raise some of our concerns regarding its validity.

Allaz and Vila (1993) show that forward contracting causes sellers to increase their output compared to a spot only market structure. In a duopoly setting, both sellers want to do this, resulting in increased total output and lower price compared to a spot market only structure; thus their market power is mitigated. Powell (1993) examines the behavior of sellers who may try to collude in both the spot and forward markets. Even in this case, the existence of a forward market puts downward pressure on the expected future spot price, so buyers will want to contract in the forward market because that will drive down spot market prices, even if buyers suspect collusion on the part of sellers. Thus once again we find that the forward market acts to put downward pressure on a spot price that is supported by market power. Green (1993) argues that price competition in the forward market will push price in that market and the spot market down to marginal cost. This would indicate a complete elimination of market power. To the extent, however, that forward market contracting is incomplete, prices in both markets will exceed marginal cost. Risk aversion on the part of buyers will increase the demand for forward contracts. Risk aversion on the part of sellers will increase their desire to sell forward contracts. Thus risk aversion propels both parties towards forward contracts, which, when signed, cause a mitigation of market power. When Green introduces asymmetric information, the balance of power shifts back to the sellers, who want to raise price in the spot market to influence expectations about price in the forward market. In this case, which is probably the most realistic one of those analyzed, the equilibrium level of forward contracting is less than complete, so that price will never fall to marginal cost. But to the extent forward contracts are adopted, their effect is still to mitigate market power.

None of these papers examine the effect of the threat of entry on these dynamics. Newbery (1995) adds this element to the analysis. His principal observation is that the low cost of new entry limits the extent to which spot and short-term contract prices can rise.

For policymakers, the important question is whether the potential for contract markets to limit market power in the spot market is adequate to produce "workable competition." This requires a definition of "workable competition," which is difficult to formulate.

To understand the mechanism through which forward contracts work to reduce market power, it is necessary to examine Cournot competition. Cournot competition occurs when a small number of firms compete by setting quantities instead of prices. This can be done in a number of ways, but in a spot market it would be done through bidding a very steep supply function in the spot price auction. It is not hard to show that steep supply functions are strategically desirable.[14] Since in the limit, a vertical supply function amounts to bidding a fixed quantity[15] a Cournot model may provide a useful tool for gaining a qualitative understanding market power in the electricity spot market.

In a Cournot equilibrium, each firm takes the other firms' strategies as given (just like in a Nash equilibrium) and then optimizes its own supply quantity. In doing so, it recognizes that if it increases supply it will lower price. Thus, its profit function is:

$$\pi = p(\bar{Q} + q)\cdot q - C(q)\,. \tag{3}$$

Notice that price is a function of the output of all other firms, $\bar{Q}$, (assumed constant) plus the output of the firm in question. If demand has an elasticity of ϵ, and if the firm's share of the market, $q/(\bar{Q}+q)$, is denoted by s, then maximizing profit results in a markup give by equation **(2)**.

To understand forward contracts, we introduce into the above Cournot model a CFD covering quantity *d*. Such a contract pays the supplier $(p_0-p)\cdot d$, where p_0 is the contract price. This gives us a new profit equation:

$$\pi = p(\bar{Q} + q)\cdot q + (p_0 - p)\cdot d - C(q)\,. \tag{4}$$

With this new profit equation, maximizing profit results in a markup is given by:

$$\frac{p-c}{p} = \frac{s-d}{\epsilon}\,. \tag{5}$$

As the coverage of supply by CFDs becomes more complete, the markup decreases until, at full coverage, markup becomes zero.

This argument does not consider repeated rounds of contracting. Although the lower markup applies only to the spot market, it undermines the price of future CFDs. Taking this into account may substantially change equation **(5)**. Consequently it is not entirely clear whether the mitigating effect of contracts on market power in the spot market will hold. None of the models examined consider this complication. Therefore we cannot be sure whether to rely on this mechanism, or to what extent it will operate. This remains a subject for further research.

---

14 By bidding a very steep, possibly vertical, supply function firm A assures the others that if they raise the market price (by cutting back on their supply) it (firm A) will not respond be stealing some of their market share. This makes it safe for other firms to cooperate in the Cournot equilibrium by cutting their quantities and raising price to the Cournot level. Thus we should expect firms to bid steep supply functions in a spot market with few suppliers.

15 A "vertical supply function," is actually horizontal at zero price up to some limiting quantity Q_o, at which point it becomes vertical. Consequently the supplier is offering to sell up to Qo at any positive price, but refusing to sell more at any price. This is equivalent to bidding Q_o.

6.1.3 *The Effect of Losses on Market Size*

Transmission network losses can be viewed approximately as normal transportation costs, even though they are proportional to the square of the power flow. We believe that network losses are generally small enough that when they are the only barrier to trade the market is generally quite competitive. For instance, when the Western Regional Transmission System is constrained only be losses, two-thousand mile trades between northern and southern regions are commonplace. Moderate market power problems may arise when these trades are blocked by congestion, effectively isolating Washington and Oregon from cheap power in the Southwest. Extreme market power problems could arise after deregulation due to local congestion, such as occurs on the path into San Francisco which is limited, for reliability reasons, to a flow that is approximately half the city's peak consumption level. The remaining power is currently being supplied by local generation, which is owned entirely by PG&E. Because losses represent a relatively benign source of market power, their effects have not yet been investigated. When it is, the large literature on spatial price competition, which focuses on the insulating effects of distance and transport costs will become highly relevant. Because we believe losses play a less important role, we focus mainly on the effects of congestion.

6.1.4 *The Effect of Congestion on Market Size*

Because market power depends crucially on the size of a firm's market share, the geographical extent of a firm's market is also of crucial importance. Unfortunately, the geography of bulk power markets is determined by congestion and this fluctuates on an hourly basis. Although there is no limit to the complexity of geographical separations that can be caused by congestion, the effects of a congested line on market power can be broken into the following two distinct cases:

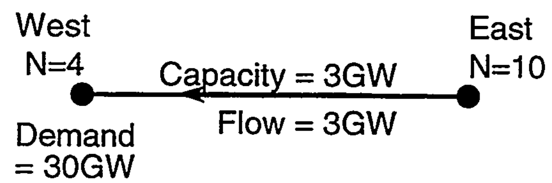
(1) Flow on the congested line simply acts as a shift in demand at each end of the line.
(2) Suppliers at each end actually "feel" competition from suppliers at the other end.

We will consider each in turn, and represent each by the simplest possible example. Each example has only two nodes and a single connecting line that is congested or is in danger of being congested. The first example has market power at only one end, while the second has market power at both.

The two cases might correspond, in the California setting, for example, to congested lines that enter the state from the North and East, and which connect it to large external markets that are essentially competitive (case 1), and to congested lines under peak load conditions that are internal to California and connect regions that have a limited numbers of suppliers, and thus exhibit market power at both ends of the lines (case 2).

Case 1

Figure 5. Congestion Supports Market Power



Because of the transmission capacity limit, market power in the west is effected very little by the 10 competitive firms in the east.

We begin with case (1), the most elementary congested-line problem. This example illustrates the most basic point about congestion and market power. In this case, the inflow of power on the congested line effectively shifts the demand function at both ends of the line. At the transmitting end, demand is increased, and at the receiving end demand is reduced. If, as in our example, there is a monopolist at the receiving end, it will mark up prices less, but it will still act as a monopolist.

In this example, illustrated by Figure 5, we call the two nodes East and West, with the East node being assumed to be competitive (10 firms). The West node represents California, and is assumed to be subject to market power on the part of suppliers (4 firms). To simplify the analysis, we represent this market power as a monopolistic supplier. We also assume for simplicity that both the monopolist and the competitive suppliers to the East have the same marginal cost.

The first step is to notice that price at the East node will equal marginal cost, while the monopolist in the West will always find it profitable to restrict demand until price exceeds marginal cost. Consequently, with price higher in the West, the line will certainly be congested with power flow from East to West. (An optimal dispatch requires this whenever there is a price difference in such a simple network.) Now the monopolist knows that there can be no response from the East to a change of supply, so the western monopolist continues to act as a monopolist facing this shifted demand. Because demand is lower, the monopolist may set a lower price, so in this sense its market power may be curbed. But the monopolist does not need to take into account any competitive behavior of firms to the East.

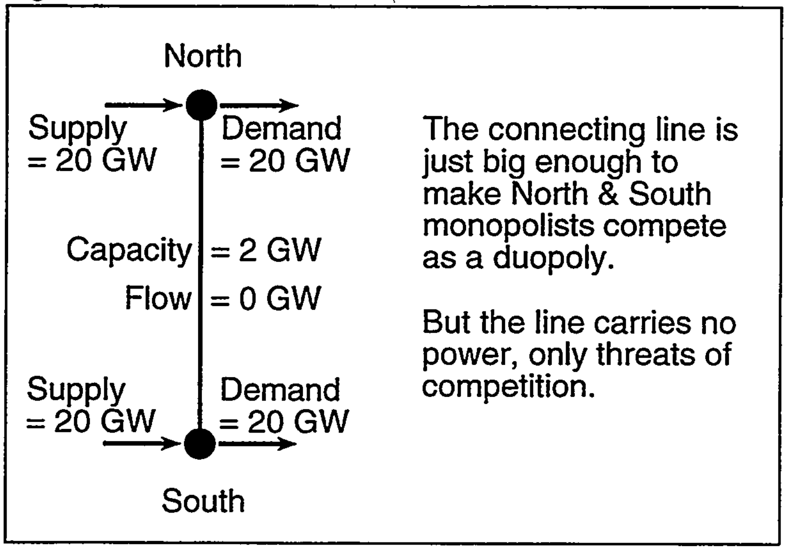
Case 2

The simplest example of case 2, closely mimics one of the main constraints in a potential California power pool. This is the constraint which occurs when transmitting power from Northern to Southern California (the so-called "south of Tesla constraint").

This north-south constraint, illustrated in Figure 6, is particularly simple because, to some reasonable degree of approximation, it breaks the market into two regions. We will make the simplifying assumption that the two regions, north and south, are identical in every respect, and that there is a monopoly supplier in each.

If these two regions were completely separate, there would be identical monopolistic solutions in the two regions. If we connected them with a very strong line, so that both suppliers could sell as much as they wanted in the other's market, there would be a duopolistic solution with lower prices. Although this result is obvious, it has one very surprising property: *No power flows on the connecting line.* This is a consequence of the complete symmetry of the problem. This means that although the line is *not* used, it is still very useful, because it keeps prices low. The threat of competition is all that is really needed, and the line (if it is big enough) provides that threat.

Figure 6. An "Under-Used" Line



- **If a connecting line is of sufficient capacity to reduce market power as much as possible, it will appear to be over built and under used.**

This raises the obvious question: how big a line is needed to induce duopolistic, instead of monopolistic, behavior? Ironically, the answer has nothing to do with actual power flow on the line, and everything to do with the threat of competition. The line sizing question is very difficult to answer, and the best we can do currently is to solve one very simple example. In this example the answer was found to be that a line big enough to carry 1/10 of the power sold at one node would be sufficient (see Borenstein, Bushnell, Kahn, and Stoft 1995 for discussion of this example). Note that this solution again depends on the unproven assumption of Cournot competition.

Probably, the most interesting problem involves a transmission line that is too small to bring about the duopolistic solution. Clearly a very weak line would provide such an example. If we imagine the two nodes connected by an extremely weak line, then both suppliers would almost ignore each other, because both would know the other could affect them very little. Thus we should have something close to the two-monopoly solution. But because of the symmetry of the problem, the line should be uncongested and this adds a duopolistic character to the solution. The actual solution is quite complex, but it involve randomness and congestion in both directions. This example allows the following conclusion.

- **A line that is on-average uncongested may still be too small to reduce market power at both ends as much as a larger line would.**

This reinforces our previous conclusion as can be seen as follows. This conclusion tells us that even if a line is seen to be uncongested on average, it may be to small to produce the maximum possible market-power-reducing effect. In order to eliminate its residual contribution to market power, its capacity will need to be expanded. When capacity has been expanded sufficiently to eliminate any contribution to market power, it will appear to be even more over built than previously.

The implication of our analysis is that in situations where generation market power is a problem, there may well be value in building a more robust transmission system than can be justified on the basis of the economics of power flows alone. Of course there are costs to overbuilding the network, so a cost-benefit analysis will be necessary.

6.1.5 *The Effect of Competition and Market Power on Reliability*

Competition will force prices down close to marginal costs. How, then, will generators cover their fixed costs? If they cannot, how will adequate capacity margins be maintained? It is useful to analyze this issue by generation market segment, distinguishing between baseload, mid-range, and peak-load capacity. Our analysis of reliability (as well as market power) indicates that the design and operation of the market during times of peak demand, or unexpected loss of supply, needs much more attention. The Staff of the Public Service Commission of Wisconsin makes a similar point in its assessment of restructuring (PSCW 1995). A preliminary back-of-the-envelope calculation indicates that a spot price as high as 50¢/kWh may be needed at peak load in order to induce the correct level of investment in peaking generators.[16]

We begin with the issue of base and mid-range capacity. First, it is true that when only base-load capacity is required, price will fall to the marginal cost of base-load generation, which can be quite low. Thus, revenues collected during these times will make no contribution to the fixed cost of base-load capacity. However, it is also true that when mid-range or peaking capacity is required, price will exceed base-load marginal cost (and this price will be paid for all generation) so base load will, during these periods, collect revenues contributing to their fixed costs. In Appendix A, we present a detailed argument which shows that if the system has the optimal mix of base-load capacity, then the base-load capacity will earn exactly what is required during high demand hours, to cover all costs and earn a normal rate of return on

---

16 This value (50¢/kWh) corresponds to an estimate presented by Michael Schnitzer at the Harvard Electricity Policy Group seminar (9/27/95) as an estimate of the energy charge that would just cover the cost of peakers in the current system. It is also the value arrived at theoretically based on an elasticity of demand of 0.2 and a fairly reasonable load-duration curve. This question deserves much more attention.

investment. Moreover, if the system has too much or too little base-load capacity, the market price will induce low or high profit levels (respectively) that will cause investors to adjust the level of base-load capacity towards the optimum ratio.

Mid-range capacity of any variety, from just above base load to just below the last peaker in the loading order, will be treated similarly by the market. Each variety of capacity can only earn enough revenue to help it cover marginal costs when capacity that is higher on the loading order is called into service. But during this time, it will earn enough to ensure that the proper amount of that type of generation is in the mix of generation technologies.

This brings us to the second issue; how will the peaking technology cover its fixed costs. Whenever this technology runs, it is by definition the marginal technology, and so one would expect a competitive market to drive price down to its marginal cost, thus preventing it from covering fixed costs. However, once the last peaking unit has been brought on line, the market no longer behaves competitively, because we are in a condition of short supply. At this point, any increase in demand will raise price above marginal cost.

The behavior of the market in this condition is very delicate. Price must be raised to clear the market, but for this to work, demand must respond to price. If demand is totally unresponsive, the market will fail to clear, which in a power market means rolling blackouts; not a happy turn of events. If demand is merely very insensitive to price, then the market will clear but at a very high price. In general this is to be considered the desired outcome, as a very high price is needed to cover the fixed costs of capacity that runs only a few hundred hours per year. Very preliminary calculations indicate that a price as high as 50¢/kWh may not be unreasonable during the annual demand peak.

This analysis highlights several problems that deserve serious attention before a competitive market is launched. These include, how to insure that the market will clear, how high a price should be considered a healthy outcome of competition, and how best to induce a large demand response to price. The good news is that if demand is, or can be made to be, quite price responsive, this response can take the place of a large amount of peaking capacity which will result in substantial savings for consumers.

6.2 Conflicts of Interest in the Vertically Integrated Firm

There are a variety of ways in which the vertically integrated firm can exercise its control over the transmission network to frustrate competition. This can occur in both the short-run trading market and the long-run capacity expansion market. In Section 6.2.1 we address the problem of long-run access. We examine whether the introduction of an independent system operator can mitigate access problems in Section 6.2.2.

6.2.1 *Long-Run Wholesale Access*

A recent competitive bid in Southern California illustrates the long-run problem of competitive wholesale access under vertical integration. In this 1993 case, San Diego Gas and Electric (SDG&E) was seeking approval from the California Public Utilities Commission (CPUC) to repower its South Bay Unit 3 from steam generation to a combined cycle configuration. After initial indications that the CPUC would approve of this project, independent generators complained. As a result, the CPUC ordered SDG&E to seek competitive bids as a "market test" of the project economics. All the independent projects which were bid were located on the periphery of the SDG&E system, relatively far from the load center and required transmission upgrades. The South Bay 3 unit is located at the load center and would provide reactive power support. To evaluate the bids, SDG&E hired a consultant. Absent the location factors, the South Bay 3 bid was not superior. To evaluate the locational effects (costs and benefits), the consultant relied on the utility's own estimates. With locational effects estimated in this way, South Bay 3 was deemed the preferred alternative. The CPUC ordered that the consultant's evaluation be made public (RCG/Hagler Bailly 1993). This was unexpected by SDG&E and the bidders, each of whom was upset, but for different reasons. The bidders were unhappy because details of their offers were discussed. The utility was subjected to criticism concerning the independence of its consultant. In the end, the utility dropped the project, citing a reluctance to make any long-term commitments in the face of increasing competition.

This set of events illustrates the information asymmetry inherent in assessing transmission costs and benefits. It is possible that SDG&E could have contracted with a consultant who had substantial expertise in transmission issues and who would have been in a position to make a truly independent assessment. Given the outcome, i.e., the extent to which the utility's competitive advantage appeared to lay in their favorable location, the utility would have a disincentive to seek out advice that may have gone against its interest. On the other hand, it isn't clear that the utility could have known in advance that locational effects would be the fundamental determinant of bid ranking. The motivation of the CPUC in forcing the disclosure of the bid evaluation report is unclear. The effect of the disclosure was to reveal the inconsistency between the treatment of locational effects in this setting with what the CPUC was requiring in the statewide competitive bid known as the Biennial Resource Plan Update (BRPU). The BRPU policy on transmission costs was intended to provide bidders with information in advance on the upgrade costs associated with siting at different points in the network. This policy would remove much of the information asymmetry between the utility and non-utility bidders in a competitive procurement. Since it focuses only on costs and does not address benefits,[17] the BRPU policy does not give optimal siting signals, although it is a considerable advance compared to previous methods.

---

[17] No approach based solely on upgrade costs can incorporate the benefits of siting to relieve congestion.

6.2.2 *Independent System Operator*

The potential development of competition in retail markets has elevated sensitivity to the conflict of interest problem in short-term markets. Nowhere has this discussion been more intense than in California, where the CPUC investigation of retail access (CPUC 1994) has stimulated substantial discussion. One important area of agreement to emerge from the California debate over retail access has been consensus on the need for an independent system operator (ISO) under any of the plans being discussed (Stalon and Woychik 1995).

There are several countries in which the ISO concept has been implemented, or is under active discussion, and where complete vertical divestiture has not been, or is not intended to be implemented. These cases are interesting because they raise the possibility of introducing more trade and competition without vertical ownership changes. The country with the longest experience of this kind is Spain (Kahn 1995). In 1985, the vertically integrated utilities sold their high voltage transmission lines to a government entity, Red Electrica de Espana (REE), which became the operator of the grid and which centrally dispatches the entire system as an integrated national pool. Utilities, both private- and government-owned, retain generation assets and distribution assets. Until quite recently, REE was effectively unregulated, and operated under ministerial control. The Spanish government created a regulatory commission in 1995, which will exercise oversight on REE, among other responsibilities.

Similar arrangements are being proposed in Alberta, Canada (ADOE 1994). In this case, there are also both private and government owned utilities. They already transact all electricity, except for self generation, through a centralized pool. The government and industry have agreed to open up this market to more competition by providing for access to the pool, initially for existing self-generation plants and, in the long run, to any new generator. Implementing open access will involve setting up an oversight function for the pool and grid operations that will involve all industry participants. Ownership transfers of transmission assets are not anticipated, nor will there be any vertical divestiture. Because Alberta already bases transactions on pool operation, the incentive for generators to sell to their affiliated distributors has already been attenuated.

Although an ISO could be expected to increase market transparency, there have been barriers to this, at least in the case of Spain. REE, as the agent of government electricity policy, dispatches the system to meet certain national objectives regarding the use of expensive locally produced coal. Preferential treatment for local coal production is a ubiquitous practice in Europe, even though it is opposed by the European Union (De Paoli and Finon 1993). The result in the Spanish case is that REE must constantly be adjusting its implicit dispatch rules to meet the coal target, given the other conditions on the system (i.e., expected hydroelectric and nuclear production, demand levels, maintenance schedules).

The arrangements in Spain and those proposed in Alberta mitigate conflicts of interest in vertically integrated utilities by interposing a strong market-making institution. The separation of ownership of transmission lines in Spain provides stronger assurances that there will be no manipulation of access than the proposed sharing arrangement in Alberta. On the other hand, the governance structure in Alberta may be able to exercise better control of the pooling and access functions than in Spain, if only because preferential access is not reserved to a special class of suppliers.

In the U.S. context, the discussions of the ISO concept have focused on the issue of its commercial role. Various pooling proposals address the conflict of interest problem by assigning the commercial function of matching supply and demand to the ISO, along the lines of central pooling institutions around the world. The argument against assigning this commercial role to the ISO is that it creates a difficult regulatory problem, and that there is an arbitrariness to such dispatching that need not produce efficient outcomes (Wu *et al.* 1994). Alternative visions of an ISO have been discussed in Section 5.3. With an ISO that does not play the role of grid merchant, there may still be a problem due to vertical relations. To achieve the benefits of decentralized trade, there should not be incentives for commercial relations that are dictated by vertical ties. Thus, in the California debate, for example, the proposal for direct access and no central pooling requires that utilities spin-off their generating assets (Knight 1995).

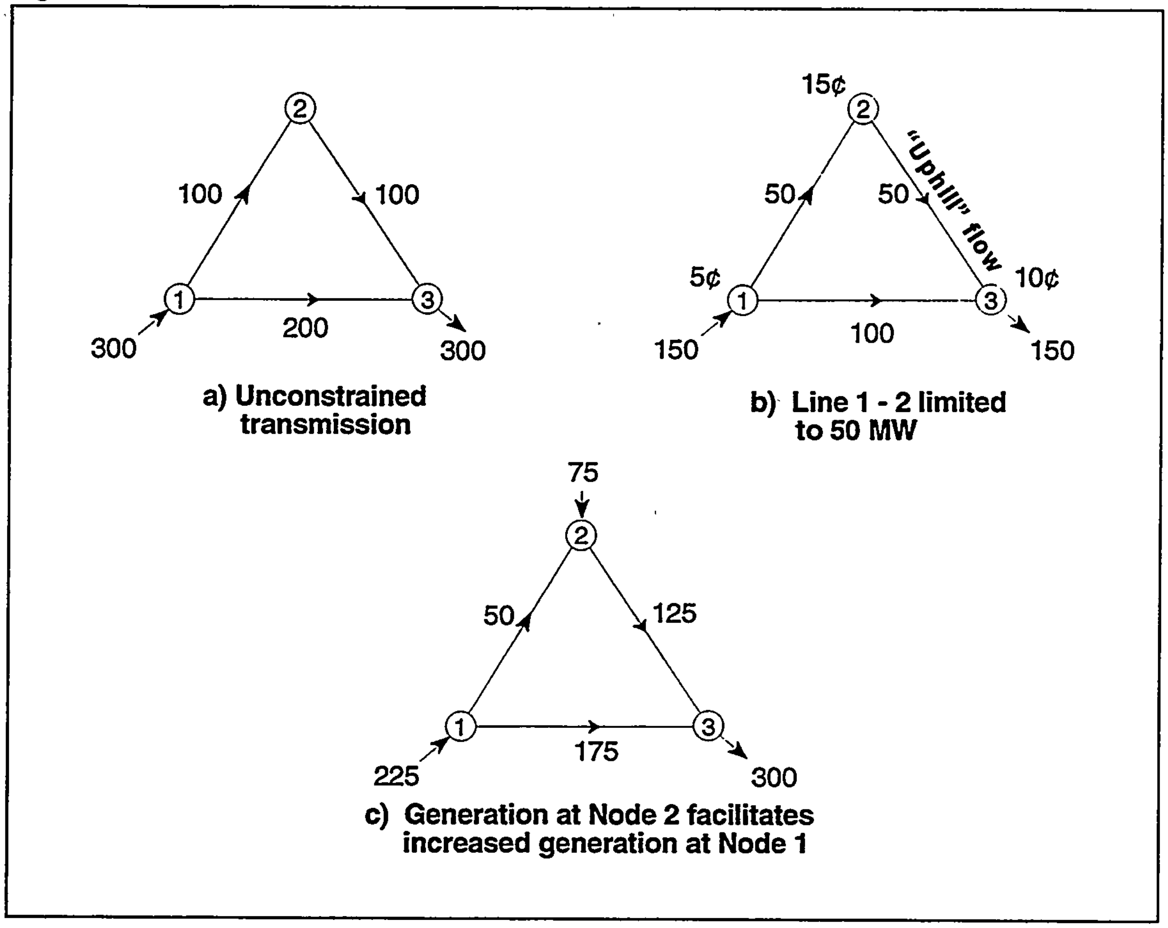
Ultimately, vertical issues must be addressed either by structural or regulatory controls. While spot market institutions, particularly a mandatory pool, may ameliorate preferential trading with affiliated entities, there may still be a problem in the contract market. Even in the UK, the regulator has had to review affiliate relations involving the investments of distribution companies in IPP projects from which they also purchase (OFFER 1992c; OFFER 1993a).

6.3 Blocking Behavior

Network topology has a strong influence on the potential for trade. It is a fact emphasized by the technologists that electrical networks do not function the way ordinary "transportation" networks function, and therefore analogies from other network industries are approximate at best. There are several ways in which particularly situated agents can block the transactions of others. Hobbs and Kelly (1992) illustrate straight-forward examples based on the network topology in New York state. An example with an international cast is given in Bjortvan and Tjotta (1993), where the position of Sweden in the Scandinavian countries effectively limits the ability of Norway to export throughout the region. The blockades examined in both cases are quite transparent. The only path for exchange lies across the territory controlled by the blocking agent. In both of these studies, the authors conclude that pricing policies which allocate a large share of the benefits of trade to the blocking agent are required to reach efficient trade levels.

Electrical properties of the network can also be used to affect trade blockades. This is usually due to the presence of weak lines in the network which limit flows. Generators located near

Figure 7. Line Constraints Produce Market Power



these lines are in a position to facilitate or block the output of other generators. Chao and Peck (1995) give a useful simple example of this phenomenon, which we reproduce below as Figure 7. They adopt the standard pedagogic convention for illustrating network constraints, namely a three node network with two supply nodes and one demand node, linked by lines of equal impedance.[18] Figure 7a shows the case where the generator at Node 1 injects 300 MW, and the generator at Node 2 does not produce at all. The powerflow equations determine that the injection at Node 1 is divided in the ratio of 2 to 1, corresponding to the relative impedances of the short path from Node 2 to the demand at Node 3 and the long path that goes from Node 1 via Node 2 to the load. Since the lines are of equal impedance, the long path has twice the impedance of the short path. Figure 7b shows the effect of a 50 MW transfer limit on the line from Node 1 to Node 2. With only the generator at Node 1 producing, this line limit reduces the total supply to only 150 MW.

---

18 Impedance is the measure of the ease with which power flows on the lines of a network.

Figure 7c shows that producing power at Node 2 will allow additional production at Node 1, while still respecting the limit on the line from Node 1 to Node 2. The flows from Node 2 will offset incremental flow from Node 1 to keep the net flow at 50 MW, as long as every incremental MW at Node 1 is matched by an incremental MW at Node 2. This gives the generator at Node 2 substantial influence over the market. His output determines how much above 150 MW the generator at Node 1 can produce. Depending on the demand elasticity at Node 3, the Node 2 generator can charge a high price for this service. If the price offered is not high enough, the generator at Node 2 can simply withdraw from the market and force production and demand back to 150 MW.

Notice also that the congestion on line 1-2, produces an "uphill flow." By this we mean that if the nodal price at nodes 1 and 3 are 5¢ and 10¢, respectively, then the optimal nodal spot price at node 2 is 15¢. The optimal price at node 2 is the net system benefit of an additional kWh. Every kWh injected at node 2 has a direct benefit of 10¢ to node 3. But every kWh injected at node 2 also has an indirect net benefit of 5¢ because it allows node 1 to ship an additional kWh to node 3. Thus, the value and price at node 2 is 15¢/kWh, and the flow from node 2 to node 3 is an "uphill" flow: it takes power from an expensive node to a cheap node. Any price above 15¢ at node 2 would involve the exercise of market power; i.e. be above the social value of power injected at that node.

This example uses the simplest possible characterization of electricity networks. Only the DC powerflow is taken into account; even line losses are neglected. In reality other network constraints also allow generators to exploit local geographic position for market power. Kahn and Baldick (1994) give an example, also in the three node, two generator, equal line impedance style, where voltage constraints coupled with one seller's refusal to support network requirements become a tool for market share gains. These examples resemble some of the documented abuses of market power by the large generators in the UK (OFFER 1992d; OFFER 1993b).

6.4 Will Comparable Service Standards Work?

In light of the potential abuses of market power that may occur, it is reasonable to ask whether these can be sufficiently mitigated by the regulatory approach currently being adopted by the FERC Mega-NOPR (FERC 1995). If not, more significant steps will need to be taken. The independent system operator concept, common to the competition models analyzed in Section 5 and also discussed in Section 6.2.2, is a more significant reform of bulk power institutions than the comparable service standards currently being implemented by FERC, which only require that the vertically integrated firm offer service to others on a nondiscriminatory basis.

The main critique of the comparable service standard approach lies in the difficulty of enforcement. The staff of the Federal Trade Commission (FTC 1995), for example, argues that the time sensitivity of electricity transactions would make compliance monitoring extremely difficult. Regulation would be required "virtually transaction by transaction." This is simply not feasible.

The counter-argument relies on reputation effects. Since regulatory oversight is necessarily incomplete, it is rational for attention to be concentrated on firms that have developed a reputation for non-compliance. Such reputations are developed by complaints from affected parties. As long as the gains from individual episodes of the exercise of market power are small, a utility might feel that the costs in terms of future oversight would be large if they were to engage in market manipulation. Only time will tell if these effects are strong.

To illustrate the enforcement problem it is useful to examine the kinds of litigation that may develop in an open access regime. We summarize a relevant case in the next section and then discuss its implications.

6.4.1 *The Cleveland Electric/ Cleveland Public Power Dispute*

The dispute between Cleveland Electric Illuminating (CEI) and Cleveland Public Power (CPP) over an emergency incident illustrates what may be an extreme, but nonetheless quite relevant, example of enforcement issues arising in a competitive regime. This dispute is discussed at length by Porter (1995), who summarizes the FERC litigation. We borrow from that review here.

CPP is a small (200 MW peak) transmission dependent municipal utility that competes with CEI for retail load in the city of Cleveland. This type of retail competition is unusual in the U.S. It is facilitated legally by the particular nature of the franchise laws in Ohio. Not only do the two utilities compete at the retail level, but CEI is required to provide transmission service to CPP. Under the terms of an anti-trust settlement associated with nuclear power plant licenses granted to CEI and its joint venturers, CEI must provide CPP with superior transmission service, not simply comparable service.

On June 16, 1994 an emergency condition occurred in the Cleveland area due to heavy summer loads and scheduled maintenance on several of CEI's plants. As a result of heavy import loadings on CEI's lines (more than 60% of area demand), voltages on the CEI-CPP interconnections declined. CEI requested that CPP reduce its reactive power demand, which would have helped restore voltage. This did not occur. Next, CEI ceased service to its own interruptible customers, and curtailed non-firm transmission service to CPP. There was still no change in deliveries to CPP. When voltage at the interconnection declined below 95% of the rated level, CEI began rolling blackouts to 38,000 of its own retail customers. After repeated requests, CPP reduced its own real power demand and voltage increased. Within two hours service was restored. Subsequent to these events, the parties filed complaints against each other at FERC.

Although the relationship between CEI and CPP is governed by contracts, there was still considerable disagreement about whether each side had met its respective responsibilities. CEI complained that CPP did not have firm transmission contracts for all of its demand and had failed to meet its long standing obligation to install capacitors or otherwise take responsibility for controlling its reactive demand. On its side, CPP argued that CEI was irresponsible for having so much capacity unavailable during the summer period, that CEI interrupted deliveries, and that CEI should have complied with requested changes in CPP schedules. The two sides also dispute the precise factual sequence of events, and the voltage standards that ought to be applicable.

This case raises questions about the enforcement of contracts when situations are complex, the parties have conflicting economic interests, and one party depends upon the other for critical services. FERC was called upon to intervene as this situation developed, but its advice did not appear to affect the course of events significantly. It is quite likely that the litigation costs arising out of this situation will exceed the direct damages caused by the emergency. It is also unclear that there is really a situation of market power abuse involved in these circumstances. There does clearly seem to be an atmosphere of mutual mistrust between the parties that is not conducive to longer term assurances that similar problems would not arise in the future.

The issues of the appropriate voltage standard and the extent to which either party was out of compliance with responsibilities do not seem to be appropriately adjudicated by FERC. It would be preferable for an industry-based organization of some kind that was closer to the facts than FERC to make such determinations. This kind of role would be more consistent with an independent system operator enforcing network standards than a federal regulator. The best treatment of problems of this kind is before the fact rather than after. Therefore, if it is the responsibility of wholesale customers to be responsible for their voltage demands, this should be enforced as a routine matter of grid management, rather than arising ex post as part of a contract dispute.

Increasing competition in the bulk power market will inevitably lead some participants to "lean on the grid" in one way or another. When such behavior arises, there are likely to be contract disputes, if not worse consequences. The FERC standard of comparable service implicitly assumes that these events will be very infrequent and minor. If they in fact interfere with the functioning of a competitive market, then a more substantial reform of wholesale market institutions will be necessary.

6.4.2. *The General Argument for an ISO*

The Cleveland dispute provides a telling example of the type of problem that can arise in a competitive environment where "open access" has been implemented without the benefit of an ISO. But one example of an open-access system with serious problems does not argue that the design of a successful open access system without an ISO will be flawed generally. We now attempt to construct a general argument to that effect.

We require three properties of a successful system:
(1) that it maintain reliability,
(2) that it allow reasonably efficient use of the grid, and
(3) that it provide nondiscriminatory access.

To meet the first of these requirements the system must ensure that demand does not exceed supply. This is a simple problem and it has three solutions:
(1) A high tariff price set ex ante,
(2) non-price rationing of demand, or
(3) a market clearing price.

The first (high price) solution would be easy to implement, but would waste grid resources. To avoid completely the need for rationing, the predetermined transmission tariff would need to be high enough to prevent excess demand under all possible market conditions. But such a high price would severely discourage the use of the grid during the rest of the year, during most of which time the marginal cost of using the grid is practically zero (no congestion and very small losses). Thus while solution 1 flawlessly solves the problems of non-discriminatory access and reliability, it fails to allow reasonably efficient use of the grid.

The rationing solution is the one that failed in Cleveland. In general, if the posted tariff is set at any level that encourages at least moderately efficient use of the grid, there will be many times when the demand for grid services exceeds the capacity of the grid to supply. In this case demand *must be rationed* to avoid catastrophic failure of the system. Again this is demonstrated vividly in the Cleveland example. **But the decisive problem is not the need for rationing, but the unavailability of a disinterested party to perform that rationing.** Without an ISO, rationing must be conducted by the owners of the grid, and in all cases these owners are also users of the grid. When rationing is needed, they cannot be expected to be unbiased, and even if they were, no one would believe it. The inevitable consequence is frequent and costly litigation.

The third solution, the use of a market clearing price for transmission services, requires a market in those services. Such a market can only be conducted by an ISO. Any attempt by the owning utility to conduct a market for the use of its lines, with itself being one of the principle purchasers, would rightly never pass the FERC guidelines for non-discriminatory tariffs.

The inescapable conclusion would seem to be that the FERC's guidelines simply cannot be met successfully without the use of an ISO. Further, it would seem apparent that although an ISO might be capable of implementing a moderate pre-determined price and non-discriminatory rationing, that this would be less than ideal. Rationing is rarely a desirable solution to the problem of allocating scarce resources, and it would seem to have little to recommend it in the present circumstances. Simplicity is its only virtue, and since the implementation of rules for non-discriminatory rationing are probably just as complex as those for pricing, rationing in this context will not be particularly simple. Consequently a *market* for transmission services run by an ISO would seem to be the obvious first choice for solving the open access problem. This is of course what is proposed both by bilateralist and by those advocating the nodal-pricing form of restructuring.

7 Transmission Capacity Expansion

This chapter focuses on incentive and institutional mechanisms that have been proposed to encourage transmission capacity expansion in a restructured and more competitive electricity industry. Currently, transmission capacity expansion problems are almost entirely left in the control of vertically integrated utilities who are providing for their own transmission needs. Consequently, these utilities are well informed about their needs and have good reason to satisfy them economically. Aside from the standard problems of rate-of-return regulation, transmission within a vertically integrated utility presents no special problems. But as soon as trade between utilities becomes important, the provision of transmission capacity becomes problematic. In a restructured power market, where the grid is a common resource to be share equally by all players, the problem is considerable.

Three broad approaches have been proposed to solving the transmission capacity expansion problem in a restructured industry: (1) rely on a regulated monopoly, (2) rely on regional transmission groups, and (3) rely on private investment guided by an artificial market mechanism such as awarding transmission congestion rights to investors. All three approaches show some potential, but also present serious unanswered questions. For the monopoly approach to work, the regulator must be able to specify clearly and objectively the criteria for evaluating transmission investments. This has not been done. For regional transmission groups to work effectively, they need a decision process that leads to optimal outcomes. Unfortunately, both majority rule and unanimity seem to have very serious drawbacks. In the next sections, we focus on the third approach in more detail.

There are two important sets of questions associated with transmission capacity expansion: cost recovery and investment incentives. The first involves the design of efficient prices that cover costs. Since transmission is a technology characterized by significant scale economies, marginal cost pricing will not recover total costs. There is a growing literature, originating principally in those countries that have restructured the electricity industry, about various ways to recover fixed investment costs (Rudnick 1994; Perez Arriaga *et al.* 1995). Of equal importance are the questions concerning the control over investments in grid expansion. Who can propose expansion projects, who can veto them, and who will pay for them?

7.1 Transmission Pricing and Network Fixed Costs

As we have seen, under one prominent version of a nodal pricing regime, the grid merchant will collect revenue roughly equal to losses plus revenue that is directly linked to the amount of congestion on the grid. Both of these sources of revenue are roughly proportional to the marginal value of improving the grid, which if the grid has been optimally designed will equal the marginal cost of improving the grid. But if the cost of providing grid services has a large fixed component, marginal costs will be far less than average cost and revenue will fall far short of covering the total cost of the grid. If the grid has been over built, then the marginal value of expansion will be even lower than marginal cost, and revenues will fall short of total cost by an even greater margin. In a bilateral or multilateral regime the situation can be even worse. For instance in the multilateral trading system described by Wu and Varaiya, the ISO collects no revenues whatsoever.

Open access controlled by an ISO poses an economic efficiency problem. If neither utilities nor IPPs have control over the network, they will not be inclined to expand it, so that network expansion and maintenance will need to be funded by the ISO. For this to happen the ISO needs a larger source of revenue than it can obtain from charges for loss and congestion. Therefore the ISO will need to impose some transmission charge in addition to the optimal loss and congestion charges. This additional transmission charge (perhaps similar to the "uplift" in the UK) may discourage trade, but that may be the price that has to be paid for neutrality of grid operation. Other possibilities for funding and motivating grid expansion are discussed in the next section.

7.2 The Control and Motivation of Grid Investment

In the long run, expanded trade will require expansion of the transmission network. This might be expected sooner rather than later, because the grid in the U.S. was not built with trade as its principal objective. With the highly decentralized nature of the U.S. transmission network, any new transmission system expansion necessarily involves the interests of many utilities. Traditional methods of dealing with these interactions were based on study groups of transmission-owning utilities. More recently, with encouragement from FERC, cooperative institutions, known as Regional Transmission Groups (RTGs), have been encouraged to develop. Although there is still relatively little experience with RTGs, the basic organizational goal is to broaden the range of economic interests participating in transmission capacity expansion discussions and decisions (Kahn 1994). This means primarily that independent power producers and transmission dependent utilities (mostly government owned) will be active members.

The problem of optimal transmission capacity expansion is quite complex. First, there are a variety of ways to expand the network to satisfy demand for access at a given point. Second, transmission capacity is subject to a number of indivisibilities. Third, the optimal expansion depends upon the future pattern of demand for access, which is difficult at best to forecast. Baldick and Kahn (1993a) give an example, based on access litigation in California, illustrating all of these features.

In any institutional structure designed to facilitate transmission capacity expansion, a critical problem is getting those parties who would benefit from additional capacity to propose projects that are both beneficial to themselves and do not damage others. Although these have been the goals of cooperative planning groups in the past, these principles have never been formalized explicitly in an economic framework. Bushnell and Stoft (1995) investigate this problem in the framework of "transmission congestion contracts" (TCCs). This notion formalizes the ideas put forward in Hogan (1992).

7.2.1 *Private Investment in a Public Network*

Even if the network were run independently to both limit market power and preserve reliability, this does not necessarily mean that investment decisions cannot be made privately. To motivate appropriate private investment two objectives must be satisfied; (1) the investor must be rewarded with the benefits of the investment, and (2) the investor must suffer any costs imposed on the system. To date, in our opinion, no one has proposed a system that perfectly satisfies these two requirements. However, two models may come reasonably close: Hogan's "contract network" model, and the Chao and Peck pool-based transmission bidding model discussed in Section 5.3.1.

Before examining these proposals, it is useful to look at a more straight-forward system. We begin by noting that one of the main rewards of expanding the network is simply to make use of it for expanded trade. In fact in the case of a vertically integrated utility, this is the only reward for network expansion and it is sufficient. In a pool, the reward is still sufficient in the case of a radial line to a single IPP. In this case leaving network expansion to the private investor works perfectly. The IPP will build the optimal line because it bears all costs and receives all benefits. However, expansion of shared portions of the grid is problematic. In this case, because the ISO controls use of the grid, an expander has no way to be sure of getting full use of the expansion it pays for.

Since the investor is not given control over his investment he must be rewarded in some comparable way. To the extent that he cannot use his investment as planned it must be congested and used by others. In this case, the other parties are being charged for the use of this congested, and thus obviously valuable resource. Thus, one possible rule for compensating an investor for his loss of use is to pay him the congestion charges that are being paid by those who are getting use of his expansion. This concept has been referred to as "link based rights" by Oren *et al.* (1995). Unfortunately, this seemingly straight-forward rule has very perverse incentive properties.

Under a system of link-based rights, it would be possible for an investor to build a weak line that creates congestion, and being congested it would of course be well utilized. For example in Figure 7a if line 1-2 did not exist, it might be possible to transmit 300 MW from node 1 to node 3. If someone then builds the weak, 50 MW, line shown in Figure 7b, this will limit transmission to 150 MW as shown. So it is a detrimental line. Nonetheless, the weak line would earn 10¢/kWh for its entire 50 MW capacity. The problem is that in a power grid, even though grid users would prefer not to ship their power on the new weak line, they cannot avoid it. Thus link-based rights, while sometimes rewarding those who make beneficial investments, can also reward those who make detrimental investments.

7.2.2 *The Investment Incentive Properties of TCCs*

A completely different system has been proposed by Hogan for his "contract network." He has proposed granting "transmission congestion contracts," (TCCs) to investors in the network. These have been introduced briefly in Section 5.2.1 above. The details of the rule for allocating TCCs to investors have been spelled out by Bushnell and Stoft (1995a) and will be summarized here.

To begin with, a TCC can be defined between any pair of nodes whether or not they are physically connected, and is defined in a particular direction, and for a particular quantity of power flow. Specifically, the TCC for q units of power from node *I* to node *j* pays its holder $(P_j - P_i)$, where P_j and P_i are the nodal spot prices. The ISO pays this to the contract holder regardless of the amount of power flowing on any particular link in the network. (Note that this definition is for a loss less network.[19]) This definition by itself tells us nothing about the incentive properties of TCCs. It is the rule for allocating TCCs to investors that determines their effect on the investment incentive.

When an investor modifies the grid, that modification can effect many parts of the grid, and can effect parts that are remote from the site of modification. Because of this we need an allocation rule that takes into account the impact of the modification on the entire network. The allocation rule does this by considering the entire network's transfer capability before and after the modification. The details of this process are complex, but we outline the most crucial facts of the allocation process.

Before the investor modifies the grid, there is an existing set of TCCs that have been previously allocated by the ISO. Since TCCs are defined by power flows between nodes, a set of TCCs corresponds to a particular dispatch of the system. By the rules of TCC allocation, this existing set of TCCs must correspond to a feasible dispatch of the grid (generation is not considered in the allocation process). Call that dispatch D0. Now if the modification is useful it will make possible a "greater" feasible dispatch. In fact, it will make possible many different feasible dispatches that are in some sense greater than D0. We can now define the allocation rule for TCCs.

Definition: An investor in the grid is rewarded with a set of TCCs corresponding to the difference between D1 and D0, where D0 is the dispatch corresponding to the set of

[19] Bushnell and Stoft (1995b) have extended the definition of TCCs to lossy networks, and have shown that all their important incentive properties transfer to the more realistic setting.

previously allocated TCCs, and D1 is a dispatch, of the investor's choosing, that is feasible after his modification of the grid.

We illustrate this rule with a simple but informative example. Assume the initial network consists of a single 1 MW line from A to B, and that someone owns a 1 MW TCC from A to B. Thus dispatch (D0) corresponding to the initial set of TCCs is 1 MW injected at A and 1 MW withdrawn at B. Now assume that an investor upgrades this line to a capacity of 2 MWs. To what will he be entitled under the allocation rule? One feasible dispatch is 2 MW from A to B. Choosing this for D1 will entitle him to a 1 MW TCC from A to B. But another feasible dispatch is 2 MW from B to A. Choosing this for D1 will entitle him to a 3 MW TCC from B to A. This is because when 3 from B to A is added to 1 from A to B, the result is 2 from B to A, which is the second feasible dispatch D1. Of course, there are an infinite number of alternative dispatches for the investor to choose from.

Despite the indeterminism of this allocation rule, it has been proven that no matter what allocation is chosen by the investor, the ISO will always have sufficient revenue from transactions at the optimal nodal spot prices to cover the resulting allocation of TCCs (Hogan 1992).[20] One may wonder what an investor faced with such a array of seemingly arbitrary choices should do. But the practical reality is not so confusing as it may at first appear. If the investor does not attempt to make money off the blunders of others, but instead assumes they will act intelligently, then an optimal strategy for selecting D1 is simply to choose the dispatch that he expects will actually occur after his network expansion is complete. This is a clear rule, and the uncertainty it holds is just the normal uncertainty faced by entrepreneurs in normal competitive markets.

7.2.3 *WSCC's Rated-System-Path Method*

The WSCC has implemented a set of rules for grid expansion that has some similarities with the TCC scheme (Walton 1993). Both systems have tradeable transmission rights, and both award these rights to investors based on the extent to which they improve the grid.

One difference between the two systems is that the WSCC only examines the effect of the investment on a limited portion of the grid, which they term the "rated system path." However, when Walton describes the philosophy of the WSCC mechanism, he says "Transmission capacity rights are determined by a project's net addition to total system capability." Because the power grid in the West is so loosely connected, the procedure of

---

20 Remember that the ISO makes money from the trade of electricity in a nodal-spot-pool system because of the nodal price differentials caused by congestion and losses. The ISO's income is referred to as the merchandising surplus and has been shown to always be sufficient to cover the payments to holders of TCCs provided they are allocated as specified, and the system is optimally dispatched.

looking only at the most affected "system path" is often a good approximation to looking at the entire system.

The principle difference between the two mechanisms is that the transmission rights awarded by the WSCC are physical rather than financial. But because they can be sold, the effect will often be the same. If the line is not congested, neither WSCC rights nor TCCs have any value or effect. If the line is congested, the WSCC rights can be sold for the value of the corresponding TCC right. However there are two crucial differences. First, current ratemaking practices limit transmission charges to embedded cost, which may be less than congestion rents. Second, the current system does not provide any way to award a physical right of negative value. Because of this the WSCC specifies that "The project sponsor must demonstrate that the harm to any party has been mitigated to the satisfaction of those affected." This is a very unsatisfactory rule for two reasons. First, there is no fair mechanism for determining the extent of harm to another party. Second, not all examples of "harm" should be compensated. In the normal process of competition, it is necessary and desirable for competitors to undercut each other's market power. In this sense they are constantly harming each other. This harm has a real and computable financial cost to the party that loses market power, but it is the way in which competition generates its socially desirable effect. It is not desirable to require compensation for this type of harm, but the WSCC mechanism provides no way to distinguish this from true negative externalities.

The advantage of the TCC mechanism is that it does have an economically sensible method for evaluating and punishing negative externalities as well as system improvement. It seems possible that the WSCC method of assigning physical rights could be improved to account for negative externalities and to include effects on the whole network. If this were done, it would probably be quite similar to the TCC mechanism. In this case the decision as to which approach to use might depend on which system imposed lower transaction costs on the trading parties and was more easily enforced by the regulator.

7.3 Conclusion

Motivating optimal investment in a transmission grid appears to be a very difficult problem, whether that investment is to be undertaken by a regulated monopoly or by private investors. The problem is complicated both by the complex externalities produced by grid modification, and by the lumpiness of investment decisions. Currently, only two mechanisms have been proposed that show any real promise of solving the problem in a decentralized fashion, and no attractive mechanisms have been proposed for regulating a monopoly. The contract network method, which rewards investors with TCCs, is currently the most complete and sophisticated system. But even this method fails to compensate investors fully for positive externalities that their network improvements may generate. Beyond that, there is the question of to what extent the distribution of TCCs will match the actual dispatch of generators who own the TCCs. If the match is not close there may still be significant problems with negative externalities being insufficiently discouraged. The WSCC approach, while definitely a move in the right direction, is probably only workable in the Western U.S., as it is currently practiced. Improvements in the system might well move it towards that TCC approach, which could make it more practical in a highly meshed setting.

Although a decentralized approach to transmission investment is attractive for its potential competitiveness and for its information properties, it may face insurmountable legal barriers. Local siting resistance to new lines, which may be far worse than for point facilities like power plants, will be very difficult for entrepreneurs to overcome. A regulator may be needed to define "need" in order to issue the certificate of convenience and necessity needed to obtain eminent domain power.

8 Summary and Future Directions

In this chaper we synthesize and summarize the previous discussion. Section 8.1 argues that an ISO will be necessary for a successful open access regime. Section 8.2 identifies key questions arising in the debate over models for organizing the bulk power market. Future directions are discussed in Section 8.3.

8.1 Ensuring "Open Access" Will Likely Require an ISO

Achieving open access in electricity markets is the objective of the FERC's Mega-NOPR. But fully competitive market models, both in the bilateral and the spot-market form, rely on an ISO to accomplish this effect, while the FERC attempts to achieve it without the benefit of an ISO. In Section 6.4 we argue that an ISO is necessary for effective competition. Open access must solve the problem of excess demand for transmission to preserve reliability; the FERC in its Mega-NOPR allows this only through a predetermined tariff and non-price rationing. Because of the complexities of transmission, non-price rationing cannot be accomplished in a non-discriminatory fashion by an interested party. This leaves the predetermined tariff as the only tool for preventing excess demand. But for this to be almost perfectly effective, as is required by reliability, the tariff must be so high as to prevent even moderately efficient use of the grid.

A crucial part of this argument is the assertion that the allocation and expansion of transmission is a complex problem. Such complexities were discussed in Section 5.1, "Spot Market Theory," Section 6.1.2, "The Effect of Congestion on Market Size," and Section 7, "Transmission Capacity Expansion."

Although both the bilateral and nodal spot pricing approaches differ fundamentally from the FERC approach because of the reliance on an ISO, it is interesting to ask which more closely resembles FERC's current approach. As is made clear in Section 5.3, the essence of the bilateral approach is to attempt to separate the market for power from the market for transmission. This is done most successfully by McGuire's approach, in which the ISO assigns a transmission charge to each bilateral trade, but has no knowledge of the energy price involved. This does not constitute a predetermined tariff because it is reset every hour based on market conditions, although it does meet FERC's requirement for a nondiscriminatory tariff and open access.

Nodal spot pricing, as described in Section 5.1, takes a different approach to open access. It recognizes that power injected at any one node affects the flow of power on every line in the entire grid. Instead of trying to track these flows and calculate $n \cdot (n-1)$ transmission prices, one for every pair of nodes, the nodal spot price simply measures the net benefit of energy injected and announces a single price at each node. This price includes transmission charges that are correct according to the actual flow over the entire grid of power injected at a single node. Thus, the nodal spot price incorporates both the energy price and the transmission price and does so optimally. But in doing this, the nodal approach completely obscures the transmission charge. Thus nodal spot pricing achieves open access and non-discrimination, but does away, once and for all, with transmission tariffs.

8.2 Summary of the Bilateral vs. Spot Market Debate

The variety of ways in which bulk power markets can be organized, coupled with limited restructuring experience, makes it difficult to reach definitive conclusions. However, based on our review, several broad principles seem to have emerged from the recent discussions and the experience both in the U.S. and other markets. These include:

(1) a general agreement that price transparency is desirable,
(2) a broad (but not universal) consensus that an independent system operator (ISO) is necessary to facilitate increased trade,
(3) an increased skepticism about vertical economies, and
(4) a consensus that market distortions should be minimized.

But stated at this broad general level, these principles are too ambiguous to provide much guidance concerning the choices lying ahead for the organization of bulk power markets. To provide a more structured assessment of the choices, we outline a stylized version of the pool versus bilateral market structure debate to examine exactly how these general principles are reflected in the different market models.

Table 6 summarizes this comparison. It shows that depending upon the market model, even general principles have different meanings, and different implications. We summarize each of the categories listed in the table below.

Table 6. Market Model Summary

	Pool	Bilateral
ISO Function	Grid Merchant	Information Broker
Price Transparency	SRMC	Index
Vertical Economies	Less Important	More Important
QF Must-Run	Distorts Pool Price	Forecloses Access for Other Transactions
Contract Performance	Financial	Physical

The ISO function involves a much broader range of actions in the Pool model than in the bilateral model. The reason is that the Pool ISO is dispatching the power system based on sellers' prices, whereas the bilateral ISO is an information broker who facilitates the trading decisions of others. This distinction has been discussed more fully in Sections 5.1 and 5.3. The different conceptions of the ISO are reflected in all of the other market model attributes in Table 6.

Price transparency means different things in the two models, because price formation differs in each. When the Grid Merchant is the central clearinghouse for dispatch of the system, the resulting prices at any network node are the short-run marginal cost (SRMC) at that node. In the case where no congestion exists, there is effectively only one node and a single market clearing price for any given period of time. In the Pool model, price variance results from the time differentiation of SRMC, not from any variance at a given time. The bilateral contract model is more compatible with price indices averaged over inherently longer time horizons than the Pool model. These indices are averages of many bilateral contract prices. The potential biases in price reporting will also differ in each model. In the Pool model, there is some arbitrariness in SRMC determination. In the bilateral model, sampling error may distort price indices.

The question of vertical economies is quite unsettled. Both competitive models inherently question the role of vertical economies. In neither case, however, is it clear that divestiture of generation will be required for unbiased functioning of bulk power trade. There appears to be less emphasis on vertical economies in the Pool model, if only because of the precedents set in other countries, where electricity restructuring along pool lines has been accompanied by vertical separation. The bilateral model also seems more consistent with a vertical structure, if only because it is closer to current U.S. industry structure and practice. The current industry trend toward utility mergers may end up raising market power questions in either of these models. The traditional arguments for vertical economies in a monopoly structure may turn out to look like access barriers in a competitive model.

While everyone opposes market distortions in principle, in practice, none of the restructuring models has addressed the inefficiencies associated with the QF "must-run" entitlement. This distortion is present in both models, but its effect is somewhat different. In the Pool model, the QF entitlement distorts the Pool price by effectively lowering it. This occurs because low cost resources that might have been inframarginal in a less constrained market are more likely to determine marginal cost when they must compete for residual demand after the QF output has been dispatched on a priority basis. In the bilateral model, while there may be price effects, there may be more noticeable effects on access. Since the QF resources have dispatch priority, they potentially foreclose access for other transactions. This may not be directly visible from today's perspective, since such foreclosure is already occurring. As the market process becomes increasingly transparent, these effects will inevitably be noticed. In either model the magnitude of these effects depends upon the relative size of the QF share of the market.

Finally, contract performance standards differ in the two models. Given that the Pool ISO is a Grid Merchant through whom all physical transactions clear, the only role for contracts is financial. Indeed, a physical performance standard, where seller must physically deliver to buyer, is incompatible with the Grid Merchant concept. In the bilateral model, on the other hand, physical performance is the essence of commercial relations. It embodies the mutual commitment of the parties to trade. Physical performance as the cornerstone of a bilateral trade market may impose some complexity on the ISO, but proponents argue that this is feasible.

This descriptive assessment does not address normative or implementation questions associated with each of the models. These are formulated in the next section as subjects for future research.

8.3 Future Directions

In this section we list a few of the more important questions raised by the prospects for a more competitive bulk power market. This list is not complete, but rather it is indicative of issues that will need resolution in the future. In each case we comment briefly on the questions, indicating some of the directions for future study.

8.3.1 *Which Model Is Most Compatible with the FERC Mega-NOPR?*

The FERC Mega-NOPR represents an important transformation of the regulation of bulk power markets, principally by addressing equal access questions and to an extent by transcending the contract path approach to transmission pricing. (To the extent that "network service" tariffs avoid the fiction of contract paths within a utility franchise area, contract paths are transcended. Since there will still be loop flow among utilities, the limitations of the contract path will persist). Despite these significant steps, however, FERC's basic approach still relies on the traditional cost of service paradigm, which is inconsistent with the degree of pricing flexibility inherent in the Pool model.

In the short run, compatibility with existing FERC practice may represent an implementation advantage for a particular market model. But given the distortions due to cost of service pricing and the contract path, in the long run isn't such compatibility a barrier to economic efficiency?

8.3.2 *Which Model Has the Largest Transactions Costs?*

The models involve different kinds of transactions costs. In particular, we can distinguish the start-up costs from the ongoing transactions costs. The Pool model may have low transactions costs once implemented, but its implementation may have high start-up costs. The opposite appears to be the case for the bilateral model.

A broad view of transactions cost would extend to the costs of managing risk. This is a large set of questions many of which involve the prospective role of an electricity futures market. This issue will get increased attention as the New York Mercantile Exchange pursues its initiative to set up such a market.

Transactions costs are important, but public policy should not seek to minimize them at the cost of other inefficiencies.

8.3.3 *How Will Reliability Be Affected by Increased Competition?*

There is some reason to believe that reliability practices, if not the actual level of reliability will have to change under increased competition. The discussion of these issues in Section 6.1.3 indicates quite a different path toward reliability than current industry practice summarized in Section 2.2. Short run operational practices could differ in the market models, with the Pool model offering a more centralized approach to reliability management. While centralization has traditionally been used to manage reliability in the electricity industry, it is not clear what the potential is for managing these problems in a more decentralized fashion. High prices during peak periods will be required to ration demand if a strictly market solution is desired. This will create problems both because such pricing is unprecedented, and because it will be difficult to distinguish it from the exercise of market power, which is itself a major concern.

8.3.4 *Which Model Has the Greater Potential for Abuse of Market Power?*

The discussion in Sections 6.1.1 and 6.1.2 addresses the origin and exercise of market power in electricity generally, but not at the level of comparing the Pool and bilateral models. The recent accelerated merger movement in the electric utility industry may raise a whole new set of market power questions. Open access may mitigate seller market power, depending upon the degree of network congestion. Mergers may increase buyer market power, again depending upon the degree of network congestion.

It is not clear which of the market models contains greater risks from the exercise of market power. This will be an increasingly important issue in the discussion over the organization of bulk power markets.

8.3.5 *Is Vertical Separation Necessary?*

It may be that the culture and information links in the vertically integrated firm will lead to inevitable distortions of wholesale trade and investment. In this case, vertical separation will be required. Such separation has been the dominant pattern internationally (see Section 3). What would trigger a policy decision of this kind in the U.S.? Would separation be more difficult in our electricity industry than it has been elsewhere?

References

Alberta Department of Energy (ADOE) 1994. "Enhancing the Alberta Advantage: A Comprehensive Approach to the Electric Industry."

Allaz, B. and J. Vila 1993. "Cournot Competition, Forward Markets and Efficiency." *Journal of Economic Theory*, 59:1:1-16.

Baldick, R. and E. Kahn 1993a. "Network Costs and the Regulation of Wholesale Competition in Electric Power." *Journal of Regulatory Economics*, 5:4:367-384.

Baldick, R. and E. Kahn 1993b. "Transmission Planning Issues in a Competitive Economic Environment." *IEEE Transactions on Power Systems,* 8:4:1497-1503.

Bjorvatn, K. and S. Tjotta 1993. "Deregulation and Common Carriage in the Nordic Power System." *The Energy Journal*, 14:4:57-74.

Bohn, R.E., M.C. Caramanis, and F.C. Schweppe, 1981. "Optimal Spot Pricing of Electricity: Theory," MIT-EL 81-008WP.

Borenstein, S., J. Bushnell, E. Kahn and S. Stoft 1995. "Market Power in California Electric Markets." University of California Energy Institute Working Paper, PWP-36.

Bushnell, J. and S. Stoft 1995a. "Transmission and Generation Investment in a Competitive Electric Power Industry." University of California Energy Institute Working Paper, PWP-30.

Bushnell, J. and S. Stoft 1995b. "Electric Grid Investment Under a Contract Network Regime." University of California Energy Institute Working Paper, PWP-34.

California Public Utilities Commission (CPUC) 1994. Order Instituting Rulemaking on the Commission's Proposed Policies Governing Restructuring California's Electric Services Industry and Order Instituting Investigation on the Commission's Proposed Policies Governing Restructuring California's Electric Services Industry and Reforming Regulation, R.94-04-031 and I.94-04-32.

Chao, H. and S. Peck 1995. "Market Mechanisms for Electric Power Transmission." Electric Power Research Institute working paper.

Comnes, G., T. Belden and E. Kahn 1995. "The Price of Electricity from Private Power Producers: Stage II Expansion of Sample and Preliminary Statistical Analysis." LBL-36054.

De Paoli, L. and D. Finon 1993. "Implications of Community Policy for the Electricity Industry." *Utilities Policy*, 3:3:209-221.

Federal Energy Regulatory Commission (FERC) 1989. "Electricity Transmission: Realities, Theory and Policy Alternatives."

Federal Energy Regulatory Commission (FERC) 1995. "Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities, Recovery of Stranded Costs by Public Utilities and Transmitting Utilities." Proposed Rulemaking and Supplemental Notice of Proposed Rulemaking, Docket Nos. RM95-08-000 and RM94-07-001.

Federal Trade Commission (FTC) 1995. Comments of the Staff of the Bureau of Economics of the Federal Trade Commission in FERC Docket Nos. RM95-08-000 and RM94-07-001.

Fitzgerald, T. and R. Hemphill 1993. "Multi-Utility Transmission Planning: Self Regulation and Reliability." Argonne National Laboratory.

Garber, D., W. Hogan and L. Ruff 1994. "An Efficient Electricity Market: Using a Pool to Support Real Competition." *The Electricity Journal*, 7:7:48-60.

Gilbert, R. and E. Kahn (ed.) 1996. *International Comparisons of Electricity Regulation.* New York: Cambridge University Press.

Gilbert, R., E. Kahn and M. White 1993. "Coordination in the Wholesale Market: Where Does it Work?" *The Electricity Journal,* 6:8:51-59.

Gilbert, R. and M. Riordan 1995. "Regulating Complementary Products: Comparative Institutional Analysis." *RAND Journal of Economics,* 26:243-256.

Graves, F., P. Carpenter, M. Ilic and A. Zobian 1993. "Pricing of Electricity Network Services to Preserve Network Security and Quality of Frequency Under Transmission Access." Response to FERC Request for Comments in its Notice of Technical Conference, Docket No. RM93-19-000.

Green, R., 1993. "Long-Term Contracts and the Electricity Spot Market." Cambridge University Working Paper.

Hobbs, B. and K. Kelly 1992. "Using Game Theory to Analyze Electric Transmission Pricing Policies in the United States." *European Journal of Operations Research,* 56:154-171.

Hogan, W. 1992. "Contract Networks for Electric Power Transmission." *Journal of Regulatory Economics,* 4:3:211-242.

Ilic, M., J. Lacelle-Melero, F. Nishimura, W. Schneler, D. Shirmohammadi, A. Crough and A. Catelli, "Short-Term Economic Energy Management in a Competitive Utility Environment." *IEEE Transactions on Power Systems,* 8:1 (1993) 198-206.

Jaccard, M. 1994. "Changing Canadian Electricity Markets and the Future Role of Government." *Energy Studies Review,* 6:2:103-126.

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

Kahn, E. 1995a. "Regulation by Simulation: The Role of Production Cost Models in Electricity Planning and Pricing." *Operations Research,* 43:3:388-398.

Kahn, E. 1995b. "The Electricity Industry in Spain." University of California Energy Institute Working Paper No. PWP-32.

Kahn, E. 1994. "Regional Transmission Groups: How Should They Be Structured and What Should They Do?" *From Regulation to Competition: New Frontiers in Electricity Markets,* ed. M.Einhorn, Kluwer Academic Publishers.

Kahn, E. and R. Baldick 1994. "Reactive Power is a Cheap Constraint." *The Energy Journal,* 15:4:191-201.

Kahn, E. and R. Gilbert 1993. "Competition and Institutional Change in U.S. Electric Power Regulation." University of California Energy Institute Working Paper No. PWP-011.

Kahn, E., C. Marnay and D. Berman 1992. "Evaluating Dispatchability Features in Competitive Bidding." *IEEE Transactions on Power Systems* 7:3:1259-1265.

Kaserman, D. and J. Mayo 1991. "The Measurement of Vertical Economies and the Efficient Structure of the Electric Utility Industry." *Journal of Industrial Economics*, 39:5:483-502.

Kirby, B., E. Hirst and J. Vancoevering 1995. Identification and Definition of Unbundled Electric Generation and Transmission Services, ORNL/CON-415.

Knight, J. 1995. "Customer Choice Through Direct Access: Charting a Sustainable Course to a Competitive Electric Services Industry." California Public Utilities Commission.

Kovacs, R. and A. Leverett 1994. "A Load Flow Based Method for Calculating Embedded, Incremental and Marginal Cost of Transmission Capacity." *IEEE Transactions on Power Systems,* 9:1.

Laffont, J. and J. Tirole 1993. *A Theory of Incentives in Procurement and Regulation.* Cambridge, Mass., MIT Press.

Le, K., R. Jackups, J. Feinstein, H. Thompson, H. Wolf, E. Stein, A. Gorski and J. Griffith 1990. "Operational Aspects of Generator Cycling." *IEEE Transactions on Power Systems*, 5:4:1194-1203.

Lee, F. and Q. Feng 1992. "Multi-Area Unit Commitment." *IEEE Transactions on Power Systems* 7:2:591-599.

Lee, R., M. Beshir, A. Finley, D. Hayes, J. Hsu, H. Peterson, G. DeShazo and W. Gerlach 1995. "Application of Static VAR Compensators for the Dynamic Performance of the Mead-Adelanto and the Mead-Phoenix Transmission Projects." *IEEE Transactions on Power Delivery* 10:1:459-466.

Levin, R. 1995a. "Centralized Pools: Description, Critique, and Recommendations: A Contribution to the Dialogue." New York Mercantile Exchange.

Levin, R. 1995b. "Economics of Forward Contracting: Focus on Electricity Market; Comparison With and Without Mandatory Pool." Paper presented at the Rutgers 1995 Annual Western Conference.

McGuire, Bart, 1995. "Price-Driven Coordination in a Lossy Power Grid," University of California Energy Institute mimeo, 30 November.

Mistr, A. 1992. "A Proposal for Fundamental Reform of Transmission Pricing." Virginia Power.

Newbery, D. 1995. "Power Markets and Market Power." *The Energy Journal*, 16:3:39-66.

New York Public Service Commission (NYPSC) 1991. Report on the Management and Operations Audit of the New York Power Pool.

Oren, S. and S. Smith 1992. "Design and Management of Curtailable Electricity Service to Reduce Annual Peaks." *Operations Research*, 40:2:213-228.

Oren, S., P. Spiller, P. Varaiya and F. Wu 1995. "Nodal Prices and Transmission Rights: A Critical Appraisal." *The Electricity Journal*, 8:3:24-35.

Papalexopoulos, A., S. Hao, E. Liu, Z. Alaywan and K. Kato 1994. "Cost/Benefit Analysis of an Optimal Power Flow: The PG&E Experience." *IEEE Transactions on Power Systems* 9:2:796-804.

Pechman, C. 1993. *Regulating Power: The Economics of Electricity in the Information Age.* Kluwer Academic Publishers.

Pollitt, M. 1994. "Technical Efficiency in Electric Power Plants." Cambridge University, Mimeo.

Porter, K. 1995. "Cleveland Electric Illuminating/ Cleveland Public Power Litigation." NREL draft.

Powell, A. 1993. "Trading Forward in an Imperfect Market: The Case of Electricity in Britain." *Economic Journal*, 10:444-453.

Public Service Commission of Wisconsin Staff (PSCW), Draft Environmental Impact Statement, Investigation on the Commission's Own Motion Into the Probable Costs and Benefits of Changing Electric Utility Company Structure and Regulation, PSC Docket No. 05-EI-114, 1995.

RCG/Hagler-Bailly, 1993. "Evaluation of Bidders in the South Bay 3 Repower Market Test." Boulder, Colorado.

Ruganis, P. "Estimation of Historical New York Power Pool Marginal Energy Costs" Rochester Gas and Electric, 1986.

Schweppe, F.C. 1978. "Power Systems '2000': Hierarchical Control Strategies," IEEE Spectrum, July, pp. 42-47.

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

Sheble, G. and G. Fahd 1994. "Unit Commitment Literature Synopsis." *IEEE Transactions on Power Systems* 9:1:128-135.

Smith, W. And M. Klein, Infrastructure Regulation: Issues and Options, World Bank discussion paper, 1994.

Southern Company Services, An Economic Study of the Optimum Reserve Margin and Associated Reliability Indices for the Southern Electric System, 1991.

Speckman, B. And S. Schleimer, "A New Index Prices the Market." *Public Utilities Fortnightly*, 133:5 (1995) 30-31.

Stalon, C. and E. Woychik, "What Model for Restructuring? The Debate in the Competitive Power Market Working Group." *The Electricity Journal*, 8:6 (1995) 63-73.

Stoll, H. *Least Cost Electric Utility Planning*. John Wiley and Sons, 1989.

Tenenbaum, B., R. Lock and J. Barker, "Electricity Privatization: Structural, Competitive and Regulatory Options." *Energy Policy*, 20:12 (1992) 1134-1160.

Vickers, J., "Competition and Regulation in Vertically Related Markets." *Review of Economic Studies*, 62:1 (1995) 1-17.

Walton, S., "Establishing Firm Transmission Rights Using a Rated System Path Method." *The Electricity Journal*, 6:8(1993) 20-33.

Western Systems Coordinating Council (WSCC), WSCC Unscheduled Flow Mitigation Plan, Salt Lake City, 1994.

White, M. "Valuing Market Trading Mechanisms: Evidence from Electricity Markets." University of California Working Paper, 1994.

Wu, F., P. Varaiya, P. Spiller and S. Oren, "Folk Theorems on Transmission Access: Proofs and Counter Examples." University of California Energy Institute Working Paper, PWP-023, 1994.

Wu, F. and P. Varaiya, "Coordinated Multilateral Trades for Electric Power Networks: Theory and Implementation." University of California Energy Institute Working Paper, PWP-031, 1995.

International Studies

Australia

Government Pricing Tribunal (GPT), Paying for Electricity An Issues Paper, Discussion Paper No.5, 1993.

National Grid Management Council (NGMC), The Structure of an Interstate Transmission Network for Eastern and Southern Australia, 1993a.

National Grid Management Council (NGMC), Regulatory Framework Issues for a National Electricity Market, 1993b.

National Grid Management Council (NGMC), Code of Conduct: Market Rules, 1995.

Outhred, H., "Principles of a Market-Based Electricity Industry and Possible Steps Toward Implementation in Australia." International Conference on Advances in Power System Control, Operation and Management, Hong Kong, 1993.

Outhred, H. and J. Kaye, "Incorporating Network Effects in a Competitive Electricity Industry: An Australian Perspective." paper presented at The Institute of Management Sciences Conference, Anchorage, Alaska, 1994.

Latin America

Bernstein, S., "Competition, Marginal Cost Tariffs and Spot Pricing in the Chilean Electric Power Sector." *Energy Policy*, 16:4 (1988) 369-377.

Covarrubias, A. and S. Maia, *Reforms and Private Participation in the Power Sector of Selected Latin American and Caribbean and Industrialized Countries*, Vol. II, Appendix A, Argentina, World Bank Latin America and the Caribbean Technical Department, Regional Studies Program, Report No. 33, 1994a.

Covarrubias, A. and S. Maia, *Reforms and Private Participation in the Power Sector of Selected Latin American and Caribbean and Industrialized Countries*, Vol. II, Appendix B, Chile, World Bank Latin America and the Caribbean Technical Department, Regional Studies Program, Report No. 33, 1994b.

Galal, A., "Chilgener." Chapter 9 and "Enersis." Chapter 10 of Galal, A., L. Jones, P. Tandon and I. Vogelsang, *The Welfare Consequences of Selling Public Enterprises*. Oxford University Press, 1994.

Perez Arriaga, I. The Organisation and Operation of the Electricity Supply Industry in Argentina, Energy Economic Engineering Ltd., 1994.

Perez Arriaga, I., F. Rubio, J. Puerta, J. Arculez and J. Marin, "Marginal Pricing of Transmission Services: An Analysis of Cost Recovery." *IEEE Transactions on Power Systems*, 10:1 (1995) 546-553.

Rudnick, H., "Transmission Open Access in Chile." Harvard Electricity Policy Group, 1994.

Norway

Hjalmarsson, L., "From Club-Regulation to Market Competition in the Scandinavian Electricity Supply Industry." *International Comparisons of Electricity Regulation*. ed. R. Gilbert and E. Kahn, Cambridge University Press, 1996.

Knivsfla, K. and L. Rud, "Markets for Electricity: Structure, Pricing and Performance." *Proceedings of the 17th Annual International Energy Conference*, vol. 1, IAEE, Stavanger, Norway, 1994.

Moen, J. "Regulation and Competiion without Privatization: Experiences from the Norwegian Electric Supply Industry." Norwegian Water Resources and Energy Administration, 1995.

Wiedswang, R., "The Deregulation of the Norwegian Power Market." Conference on International Comparisons of Electricity Regulation, Toulouse, 1993.

United Kingdom

Armstrong, M., S. Cowan, and J. Vickers, "Electricity." chapter 9 in *Regulatory Reform: Economic Analysis and British Experience*. MIT Press, 1994, 279-322.

Green, R. and D. Newbery, "Competition in the British Electricity Spot Market." *Journal of Political Economy* 100:5 (1992) 929-953.

Newbery, D., "The Impact of Sulfur Limits on Fuel Demand and Electricity Prices in Britain." *The Energy Journal,* 15:3 (1994) 19-41.

Newbery, D. and R. Green, "Regulation, Public Ownership and Privatisation of the English Electricity Industry." R. Gilbert and E. Kahn (ed.) *International Comparisons of Electricity Regulation*. New York: Cambridge University Press, 1996.

Office of Electricity Regulation (OFFER), Report on Pool Price Inquiry, 1991.

Office of Electricity Regulation (OFFER), Report on Gas Turbine Plant, 1992a.

Office of Electricity Regulation (OFFER), Review of Pool Prices, 1992b.

Office of Electricity Regulation (OFFER), Review of Economic Purchasing, 1992c.

Office of Electricity Regulation (OFFER), Report on Constrained-On Plant, 1992d.

Office of Electricity Regulation (OFFER), Review of Economic Purchasing: Further Statement, 1993a.

Office of Electricity Regulation (OFFER), Pool Price Statement, 1993b.

Office of Electricity Regulation (OFFER), Decision on a Monopolies and Merger Commission Reference, 1994.

Yarrow, G., "Power Sector Reform in Europe with Special Reference to Britain and Norway." American Economic Association, 1994.

Appendix: The Effect of Competition and Market Power on Reliability

There is an important concern that full competition may compete away excess profit and with it the reserve margins that provide us with reliability. If such concern is well founded then perhaps market power is a necessary evil in a competitive system, for market power could provide the profits needed to finance adequate reserve margins. Perhaps this is the case in the present marketplace, but it is not the necessary outcome of competition, and will be seen once we examine the underlying mechanisms.

We will formulate the reliability problem using the standard concepts of short-run and long-run marginal costs and long-run average costs and attempt to illuminate the roles they will play in a competitive generation market.

First let us acknowledge that setting prices equal to short-run marginal costs will not come close to covering long-run average cost.[21] This occurs because long-run average cost includes both short-run marginal cost and the fixed cost of constructing the generating facility. Since without market power firms are generally forced to set price equal to marginal cost, it would seem the market power would be necessary for their survival. In a very narrow sense this will turn out to be true. If this were broadly true then with current excess capacity, price would fall to marginal costs, and stay there until enough firms went out of business to confer on the remaining firms the market power needed to lift price high enough to cover long-run average cost. This would result in the elimination of all reserve margin. In fact, by today's measurement procedures, reserve margin would certainly become negative. This would force the system into a state of shortage, which is the only state in which generators can exercise market power, for enough of the year for them to recover their fixed costs. This is not the hoped for competitive scenario.

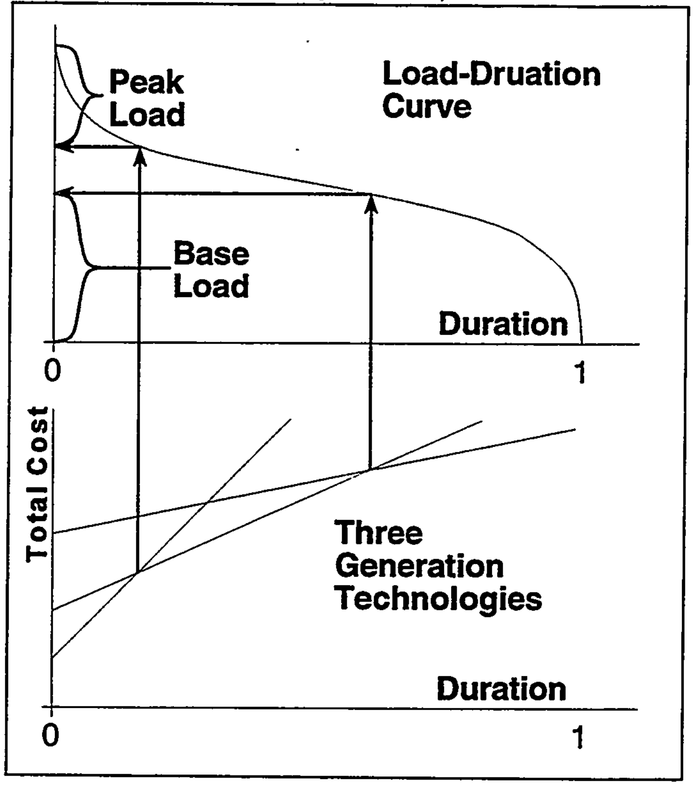
This competition-reliability conundrum is not particular to nodal pricing regimes or bilateral trading regimes. Each of these regimes, if they function as their advocates claim, reach the same set of perfectly competitive prices. Fortunately the broadest part of the solution to the paradox lies not in any particular market structure or regulation, but in the very nature of the generation industry itself.

---

21 We will not need to concern ourselves with the distinction between short- and long-run marginal costs, and the reader will be well served to imagine simply the short-run whenever we mention marginal costs. In equilibrium short-run marginal cost equals long-run marginal cost, unless the equilibrium involves holding excess capacity for strategic reasons. With excess capacity, short-run marginal costs are lower, and with capacity below the equilibrium level, short-run marginal costs are higher than long-run marginal costs.

To see this we must first broaden the problem. Reliability, properly considered, is simply one end of a spectrum of similar problems. Regarding generation, reliability is the problem of putting in place the correct number of peaking units. After that comes the problem of putting in place the correct number of mid-range units, and finally of base-load units.

Figure 8. The "Screening Curve" Method



This is an old problem, formerly solved in its simplest form by use of the "screening curve diagram" (shown in Figure 8). Of course this solution is still correct in the new competitive marketplace, but there is no longer any organization with the authority to implement this outcome. Instead a rather disorganized and dispersed market must somehow reproduce this outcome simply by relying on each firm's profit motive. Let us examine how this would work.

We start at the opposite end of the spectrum from where we first encountered our problem, and consider the plight of the base-load plant. As is well known base-load plants have the highest ratio of capital cost to output capacity, and as a consequence the lowest short-run marginal cost of operation. Thus when competition drives price down to short-run marginal cost, they will suffer most. If all plants were base-load plants, they would indeed be in serious trouble and would soon begin going out of business. This would be the first step towards re-establishing the correct generation mix, for their replacements would be mid-range or peaking facilities, but we are getting ahead of the story. In fact base-load plants will not fare so badly because much of the generating capacity has higher short-run marginal costs.

When demand increases beyond the ability of base-load generation, then the loading order moves on to higher-cost plants, and these now set the competitive price. Now the mid-range plants are barely covering operating costs, but the base-load plants are doing somewhat better and beginning to cover their fixed costs. As demand increases further, so does price, and base-load earns even higher profits. Yet each new price level is being set at exactly the marginal cost of the last plant called into service. If the profits earned by base-load plants during the non-base-load hours exactly cover the costs of the base load plants, then the generation mix, as it applies to the base, is exactly right. This is not obvious, but it can be shown by comparing the screening curve procedure with the conditions for market equilibrium. While these calculations are not inherently difficult, they are, nonetheless, beyond the scope of this report.

Mid-range plants are similar. They make nothing when only base-load is running. They make nothing (above cost) when only they are running, but when peakers are running they come out ahead. During this period they must cover their fixed costs. If they do not, then some will go out of business, and this will lengthen the duration of use of peakers and make all remaining mid-range plants more profitable. When equilibrium is established we will have the optimal number of mid-range plants.

This logic continues through, the upper mid-range, and the lower peakers, and the standard peakers, until we get to the ultimate peaker; the last plant in the loading order. This is the one and only plant (or there may be a very few identical plants) to which our original conundrum applies. How is this penultimate peaker to earn any money in a perfectly competitive market? Whenever it runs, which is rarely, it is paid only marginal cost. So its fixed costs, small as they may be, are never covered. The answer here finally seems to be market power. This penultimate plant has a monopoly during the few hours a year that it is needed, and during this time it can charge what it wants; well almost.

Beyond our penultimate peaker lies the real supplier of last resort: load shedding. In a market system, shed load should be purchased. Notice that, appropriate to its spot in the loading order, it has the highest short-run marginal cost and the lowest capital cost of all generation facilities.

From the above discussion it is clear that a competitive market can in fact provide the optimal level of reliability. There is nothing inherent in the process of bringing price down to the level of marginal cost that will prevent investment in sufficient generation capacity. However, that fact that no theoretical barrier to market-provided reliability exists does not prove that real-world markets will be capable of theoretical levels of performance. For the market to function correctly, prices must be set correctly, and customers must be able to respond to those prices within the require time frames.

The sticking point with regard to electricity pricing, is the ability of customers to receive and respond to prices quickly enough to handle system failures and anomalous demand spikes. Both occur quickly enough that some automatic load-shedding arrangements are required. If these cannot be implemented widely enough, then the cost of load-shedding will be higher than necessary and peakers will be able to exercise too much monopoly power. Alternatively, there may be system outages exactly when the price of power should be highest, which is when the peakers should be covering their fixed costs. This is a market failure that could result in insufficient reliability. Thus, both with regard to market power and to reliability, the design and operation of the market during times of peak demand, or unexpected loss of supply, needs much more attention. This point has been made by the Staff of the Public Service Commission of Wisconsin in its assessment of restructuring (PSCW, 1995). A preliminary back-of-the-envelope calculation indicates that a spot price as high as 50¢/kWh may be needed at peak load in order to induce the correct level of investment in peaking generators.[22]

---

22 This value (50¢/kWh) corresponds to an estimate presented by Michael Schnitzer at the Harvard Electricity Policy Group seminar (9/27/95) as an estimate of the energy charge that would just cover the cost of peakers in the current system. It is also the value arrived at theoretically based on an elasticity of demand of 0.2 and fairly reasonable load-duration curve. This question deserves much more attention.