THE ROLE OF DEMAND RESPONSE IN ELECTRIC POWER MARKET DESIGN

Prepared by
Steven D. Braithwait
Kelly Eakin
Laurits R. Christensen Associates, Inc.
4610 University Avenue
Madison, WI 53705-2164
608-231-2266

Prepared for
Edison Electric Institute
701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000
www.eei.org

OCTOBER 2002

Steven D. Braithwait, Ph.D.

Steven D. Braithwait, Ph.D. (University of California-Santa Barbara, 1976) is a Vice President. He specializes in competitive retail pricing with an emphasis on the measurement of customer price response. Dr. Braithwait has also worked on the design of market-based service offerings including real-time pricing and critical peak time-of-use programs. To support profitable offerings, he has developed and applied methods for wholesale market price forecasting as well as sales and load forecasting. He has provided expert testimony before public service commissions and arbitration panels in the areas of load forecasting and least-cost planning guidelines. Prior to joining Christensen Associates, Dr. Braithwait managed numerous projects in load forecasting, demand-side management and planning at EPRI. He has delivered papers at numerous industry conferences on the topics of load forecasting, customer demand management, and competitive pricing.

B. Kelly Eakin, Ph.D.

B. Kelly Eakin, Ph.D. (University of North Carolina-Chapel Hill, 1986) is a Vice President at Christensen Associates. Dr. Eakin is a specialist in the economic and financial aspects of competitive product pricing. In addition, he has experience with the organization and regulation of industries, and environmental economics. His major projects in the energy industry have included the development of innovative pricing and service designs, the assessment of customer price responsiveness and product choice, and analyses of competitive impacts of restructuring proposals. Dr. Eakin joined Christensen Associates from the U.S. Department of Agriculture (USDA) where he developed expertise in environmental and resource economics. Prior to his service with USDA, Dr. Eakin was on the faculty at the University of Oregon for seven years, where he was active in the graduate programs for the Economics Department and the College of Business. His scholarly writings have been published in a number of prestigious journals including The Review of Economics and Statistics, Journal of Human Resources, and Southern Economic Journal. Dr. Eakin also has co-edited two books, Pricing in Competitive Electricity Markets and Electricity Pricing in Transition.

Laurits R. Christensen Associates, Inc.

Laurits R. Christensen Associates, Inc. (Christensen Associates) offers a multi-disciplinary team of economists, engineers, and market research specialists, which has been providing consulting services to electric power providers for over 25 years in the areas of market restructuring, retail pricing design and evaluation, load and price forecasting, power engineering and network economics, and software design. Our team possesses institutional knowledge of numerous domestic and international industries including electric power, natural gas, health care, telecommunications, transportation, and postal services.

Edison Electric Institute (EEI) is the association of United States shareholder-owned electric companies, international affiliates and industry associates worldwide. In 2001, our U.S. members served more than 90 percent of the ultimate customers in the shareholder-owned segment of the industry, and nearly 70 percent of all electric utility ultimate customers in the nation. They generated almost 70 percent of the electricity generated by U.S. electric utilities.

Organized in 1933 and incorporated in 1970, EEI works closely with its members, representing their interests and advocating equitable policies in legislative and regulatory arenas. In its leadership role, the Institute provides authoritative analysis and critical industry data to its members, Congress, government agencies, the financial community and other influential audiences. EEI provides forums for member company representatives to discuss issues and strategies to advance the industry and to ensure a competitive position in a changing marketplace.

EEI's mission is to ensure members' success in a new competitive environment by:

- Advocating Public Policy
- Expanding Market Opportunities
- Providing Strategic Business Information

For more information on EEI programs and activities, products and services, or membership, visit our Web site at www.eei.org.

Published by:
Edison Electric Institute
701 Pennsylvania Avenue, NW
Washington, D.C. 20004-2696
Phone: 202-508-5000
Web site: www.eei.org

CONTENTS

EXECUTIVE SUMMARY

A major structural problem—a disconnection between wholesale and retail energy markets—has become apparent during the transition to competitive wholesale power markets. Wholesale price spikes have occurred on occasion as a result of a confluence of factors, including: unexpectedly high demand levels, little if any price responsive loads, short-term capacity shortages, and, some argue, market power on the part of generators. In contrast, nearly all retail customers face prices that are fixed across long periods of time, so they see no incentive to reduce usage during infrequent periods of high wholesale prices. This lack of price responsive load, or *demand response*, robs the wholesale market of a natural mechanism for relieving temporary pressures on prices, thus exacerbating the price spike problem.

Retail demand response to wholesale market conditions, where it occurs, has several important benefits. In particular, it can relieve generation and transmission constraints, reduce the severity of wholesale price spikes, reduce potential market power on the part of generators, and lead to lower overall energy prices to all consumers. These benefits are achievable, however, only if markets are designed in a way that allows wholesale market information to reach consumers directly (e.g., through dynamic, time-varying retail prices, such as real-time pricing), or if consumers are able to express their willingness-to-pay for services in a manner that can reach the wholesale market (e.g., through load reduction programs in which customers offer to reduce consumption during certain hours in return for a financial payment).

The objective of this report is to outline certain key market principles that should guide the design and selection of demand response mechanisms, and to recommend appropriate features of such mechanisms. The report first lists a number of key issues and questions that have been raised about the role of demand response in market design. It then lays a foundation for discussion of these issues by describing certain fundamental features of electric power markets. Three generic categories of demand response mechanisms are described, followed by a summary of available evidence on the recent performance of specific types of demand response mechanisms. Finally, the report describes key elements of the appropriate role of demand response in market design. A competitive market benchmark is used in deriving market design recommendations. These are modified by the practical realities of continued regulation in retail power markets.

The recent FERC working paper on standardized transmission service and wholesale electric market design and the Standard Market Design NOPR make it clear that the issue is not *whether* demand response should play a role in market design, but *how* to incorporate demand response into the standard market design on an equal footing with generation resources in order to achieve effective market performance. A number of key questions have been raised about how best to do so. Some of the fundamental questions are the following:

- How should demand response be accommodated in wholesale energy, ancillary services, and transmission congestion markets?

- What is the appropriate role of the RTO/ISO with respect to demand response—to establish rules, facilitate markets, and play the role of lifeguard/policeman, or to directly recruit demand response resources through programs such as those offered in New York, New England, and PJM?

- What is the appropriate role of both existing utilities and non-utility load-serving entities (LSEs) with respect to demand response?

- Which regulatory jurisdictional questions need to be resolved?

Is there a minimum or optimum amount of demand response? If so, how should it be determined? And which customer markets should be targeted?

The report groups demand response mechanisms into three generic *categories—dynamic pricing, interruptible and voluntary load reductions,* and *customer provision of ancillary services.* We focus primarily on markets for energy, rather than ancillary services, and draw distinctions among three types of the second category of load reduction programs: traditional *load management* programs, utility *energy buy-back programs,* and ISO/RTO sponsored *demand bidding* programs.

Many recent demand response proposals focus on the demand bidding approach, in which retail load “participates” directly in the wholesale market through a mechanism allowing customers to bid load reductions at specific prices or offer to reduce load in return for a payment offered by a supplier or system operator. The dynamic pricing approach, in which some customers choose the opportunity to face retail prices that reflect changing wholesale costs, and then decide how much electricity to use at various price levels, often appears to be viewed as only a long-term option. A competitive market view, however, suggests that dynamic pricing offers the natural benchmark for demand response mechanisms.

Conclusions and recommendations

The report’s conclusions may be summarized in a few key points.

1. Demand response is a key element of market design. The near-universal sentiment that encouraging demand response, or price-responsive load, on the part of retail customers is a necessary element of effective wholesale power market design is undeniable. The current lack of demand response leads to a number of problems in otherwise competitive wholesale markets, including wholesale price spikes, reliability problems, possible increased opportunity to exercise market power, and a perceived need for excessive reserve capacity. Two logical categories of demand response in energy markets are *dynamic pricing* and some form of *load reduction* program or *demand bidding.*

2. Utilities and LSEs are positioned to provide the natural market mechanism for demand response. In a "normal" market environment of competitive wholesale and retail markets, demand response mechanisms should arise naturally, without a need for non-market incentives, or subsidies. Some customers would willingly choose dynamic retail prices that vary with wholesale costs (especially if financial risk management mechanisms are available to limit their price risk), and LSEs would have incentives to offer such price structures. These customers would gain access to inexpensive power during frequent periods of low wholesale costs, while having a financial incentive to reduce load during occasional periods of high wholesale costs. Their LSEs would in turn have an incentive to bid price-responsive loads into the wholesale market, thus providing a mechanism for incorporating demand response into the wholesale market.

3. The dynamic pricing category of demand response mechanisms follows most naturally from a competitive market framework. Achieving a greater penetration of dynamic pricing in the face of continued retail regulation will require encouragement from state regulators for utilities to expand their standard tariffs to offer customers a choice from a range of price structures, including some that provide *dynamic pricing*. Available pricing strategies (e.g., two-part RTP for large commercial and industrial customers, and "critical price" TOU for residential and small commercial customers) can give customers dynamic pricing incentives, while still meeting regulatory concerns about revenue recovery and price risk. These retail-pricing programs provide a natural link between the wholesale and retail markets, and by acting in place of a competitive retail market allow an efficient pricing mechanism for markets in transition.

Dynamic pricing carries certain advantages relative to load reduction, or demand-bidding programs, including the following:

- Customers pay for what they consume, at prices tied directly to wholesale market costs; there is no difficult issue of measuring load reductions from potentially inaccurate baseline loads that are also subject to gaming on the part of customers.

- The demand response is provided through the customers' regular contract with their energy provider.

- Customers can focus on their primary activity of producing a product or providing a service, rather than having to act like a generator and decide whether and when to sell energy into wholesale power markets.

While individual customer response to dynamic pricing may be considered "passive," the total amount of load response by dynamic pricing customers can be forecast by their utility or LSE, who will have an incentive to do so in order to schedule load accurately in the wholesale energy market.

4. The other category of demand response mechanisms—load reduction, or demand bidding programs—faces a number of design challenges. The current environment of largely regulated retail prices and little dynamic pricing arguably creates an apparent need for ISO/RTO market intervention to encourage some form of demand response. The approach being followed in PJM, New York, and New England is *demand bidding programs* offered through the ISO/RTO, in which customers or load aggregators bid load reductions into daily market auctions along with generators. Several key questions arise in the design of such programs. Among these questions are the following, with responses:

- *What payments or incentives should be paid for load reductions?* A review of the natural economic incentives provided by competitive markets suggests that suppliers should be willing to pay no more than the wholesale cost of power for load reductions from a baseline load *for which customers have paid the retail energy price.*[1] Any higher payment implies the need for some form of subsidy to cover the difference between the revenue paid out in the incentive payment, and the cost saved by the load reduction. Naturally, the larger the incentive payment, the larger will be customers' demand response; however, the market price provides the indicator of the appropriate value and amount of load reduction.

- *How are load reductions measured (i.e., how is the baseline load calculated)?* No single best approach exists to accurately calculate each customer's baseline load during a demand response period, and avoid issues of customers "gaming" their baseline load. A competitive market approach suggests the use of pre-determined baseline loads along the lines of forward contracts for fixed quantities, which is analogous to the approach used in two-part real-time pricing.

5. The optimal amount of demand response should be determined by the market rather than specified a priori. This is a corollary to the previous conclusion. If market prices signal the value of power, then customers' response to those prices reflects the appropriate amount of demand response—low levels at low prices, and high levels at high prices. Possible qualifications to this conclusion are potential barriers to some customers of having the opportunity of choosing dynamic pricing, such as that provided by the lack of advanced metering.

6. Market designers should not ignore the benefits from the mirror image of load reductions at high wholesale prices—load increases during frequent periods of low wholesale prices. This recommendation, which receives little attention in most discussions of demand response, may provide the source of substantial benefits from improved market efficiency. However, the benefits from demand response to low prices will likely only be achieved through dynamic pricing, since demand bidding and load management programs are typically designed to operate only during infrequent periods of high wholesale prices.

---

1 This is algebraically equivalent to a condition that the supplier should be willing to pay no more than the difference between the wholesale cost and the contract retail price.

1. INTRODUCTION

Today's U.S. electric power markets stand in transition. Wholesale markets, which have been essentially deregulated, await standard market design rules from the Federal Energy Regulatory Commission (FERC). Retail market restructuring, in contrast, varies considerably by state. Customers in a few states have been given access to choice among competing energy suppliers offering different price structures. In most other states, retail prices remain regulated, and customers continue to be served by traditional integrated utilities. The problems in California have apparently slowed the retail deregulation movement in a number of states, while Texas is moving ahead.

One major structural problem has become apparent during this transition period—a disconnection between wholesale and retail energy markets. Competition in the wholesale market has kept wholesale prices in check generally. However, price spikes have occurred on occasion as a result of a confluence of factors, including unexpectedly high demand levels, little if any price responsive load, short-term capacity shortages, and, some argue, market power on the part of generators.[2] In contrast, nearly all retail customers face prices that are fixed across long periods of time, so they see no incentive to reduce usage during infrequent periods of high wholesale prices.[3] This lack of price responsive load, or demand response, robs the wholesale market of a natural mechanism for relieving temporary pressures on prices, thus exacerbating the price spike problem.

Retail demand response to wholesale market conditions, where it occurs, has several important benefits. In particular, it can relieve generation and transmission constraints, reduce the severity of wholesale price spikes, reduce potential market power on the part of generators, and lead to lower overall energy prices to all consumers. These benefits are achievable, however, only if markets are designed in a way that allows wholesale price signals to reach consumers directly (e.g., through time-varying retail prices such as real-time pricing), or if consumers are able to express their willingness-to-pay for services in a manner that can reach the wholesale market (e.g., a load reduction program in which customers offer to reduce consumption during certain hours in return for a financial payment).

Many recent demand response proposals focus on the latter approach, in which retail load "participates" directly in the wholesale market through a bidding mechanism in which customers bid load reductions into the wholesale market at given prices or offer to reduce load at a price offered by a supplier or system operator. The former approach, in which some customers choose the opportunity to face retail prices that reflect changing wholesale prices, and then decide how much electricity to use at various price levels, appears to be viewed as a long-term option.

---

2 The extreme example is provided by the California energy crisis of 2000–2001, in which wholesale prices remained high for extended periods of time.

3 Traditional utility load management programs, such as direct load control and interruptible load programs, typically provide rate discounts to customers in return for an agreement to reduce load on a mandatory basis during infrequent periods of system emergency or high wholesale power costs.

The FERC is considering mechanisms for increasing the extent of demand response in both wholesale energy markets (e.g., day-ahead, hour-ahead and real-time), as well as markets for certain ancillary services such as different categories of reserves. In an earlier EEI report, Hirst and Kirby (2001) discussed the basic features of wholesale power markets, summarized a number of existing demand response programs and the potential roles of various providers of such programs, and the barriers to greater application of demand response. Ruff (2002) reviews the economic principles of demand response and their implications for electricity markets. Several studies have attempted to quantify the beneficial effects of demand response, including its effects on wholesale market prices. Caves, et al. (2000) examined the summer 1999 price spikes in the Midwest, while Braithwait and Faruqui (2001) analyzed the summer 2000 California PX prices. Both studies illustrated that modest degrees of demand response can have substantial effects on wholesale prices at times of system constraints.

The objective of this report is to outline certain key market principles that should guide the design and selection of demand response mechanisms and to recommend appropriate features of such mechanisms. Following this introduction, Section 2 lists a number of key issues and questions that have been raised about the role of demand response in market design. Section 3 lays a foundation for discussion of these issues by describing certain fundamental features of electric power markets. Section 4 adds to the structure by describing three generic categories of demand response mechanisms. Section 5 summarizes available information on the recent performance of different categories of demand response. Section 6 provides the core discussion of the appropriate role of demand response in market design. Finally, Section 7 offers conclusions and recommendations.

2. DEMAND RESPONSE ISSUES AND QUESTIONS

FERC has made it clear that demand response should play an important role, on equal footing with supply-side resources, in wholesale electricity markets. The following list of statements is drawn from the FERC working paper on standardized transmission service and wholesale electric market design (FERC 2002a).

> "Demand response is essential in competitive markets to assure the efficient interaction of supply and demand, as a check on supplier and locational market power, and as an opportunity for choice by wholesale and end-use customers." (pp. 6–7)
>
> "Energy and transmission markets must accommodate and expand customer choice." (p. 6)
>
> "Market rules must be technology and fuel neutral. . . . Demand resources and intermittent supply resources should be able to participate fully in energy, ancillary services and capacity markets." (p. 6)

> "Demand must have the opportunity to supply operating reserves if it meets the necessary operational requirements." (p. 19)

> "Demand-side supply of operating reserves must have non-discriminatory bidding opportunities into the market." (p. 19)

> "Demand response options should be available so that end users can respond to price signals and reduce loads as they feel the price exceeds their individual willingness to pay for delivered electricity." (p. 13)

> "A bid cap, as a proxy for demand bidding, must be in effect until sufficient demand response develops in the relevant wholesale power market. . . . As a region develops substantial price-responsive demand, mitigation rules can be reduced correspondingly." (p. 23)

Thus, the issue is not *whether* demand response should play a role in market design, but *how* to incorporate demand response into the standard market design in order to achieve effective market performance.

2.1 Goals and objectives

The basic goal of incorporating demand response in market design is to improve market performance. Generally, it is hoped that the demand response function facilitated by market design rules will

- improve system reliability,
- reduce price volatility,
- increase overall economic efficiency, and
- reduce average energy prices to all consumers.

The disconnection between wholesale and retail electricity markets has led to less than optimal performance in each of these areas. A reasonable set of demand response objectives is the following:

- to provide a connection between the wholesale and retail markets,
- to make aggregate customer demands at least somewhat price responsive, and
- to capture some of the currently foregone benefits from the frequent divergence between fixed retail prices and varying wholesale costs.

2.2 Fundamental questions

Given the generally accepted need to incorporate demand response in market design, a number of key questions have been raised about how best to do so. Some of the fundamental questions are the following:

- How should demand response be accommodated in wholesale energy, ancillary services, and transmission congestion markets?
- What is the appropriate role of the RTO/ISO with respect to demand response? Is it to establish rules, facilitate markets, and play the role of lifeguard/policeman? Or should it directly recruit demand response resources through programs such as those offered in New York, New England, and PJM?
- What is the appropriate role of both existing utilities and non-utility load-serving entities (LSEs) with respect to demand response?
- Which regulatory jurisdictional questions need to be resolved?
- Is there a minimum or optimum amount of demand response? If so, how should it be determined? And which customer markets should be targeted?

2.3 Challenges and basic design issues

In addition to the above fundamental questions, a number of more detailed design issues and challenges need to be addressed. These include the following:

- How can demand response be incorporated into wholesale market design when the retail markets that are the source of demand response are in various states of restructuring, including continuing regulation, transition toward retail competition, and near complete deregulation?
- How can demand response be incorporated into market design in a manner that is fair to all parties, where fairness requires that costs be borne by those market participants that cause the costs? Fairness may also require that all market participants have non-discriminatory access to participate in demand response programs and their benefits.
- How can demand response be incorporated into market design in an economically efficient manner, without cross-subsidies?
- How can demand response mechanisms be coordinated with existing utility load management programs?

What enhancements to technological infrastructures (e.g., metering, communications) are needed to support demand response, and how will the associated costs be recovered?

These fundamental questions and design issues are addressed in the following sections, particularly Section 6.

3. FUNDAMENTAL FEATURES OF POWER MARKETS

Any discussion of the need for and role of demand response requires an understanding of certain basic facts about the unique nature of electricity and power markets. These facts have to do with the nature of electricity costs, the traditional way that regulated electricity prices have been set, how they would logically be set in a more competitive environment, and the types of demand response programs that utilities have traditionally offered.

3.1 Electricity costs

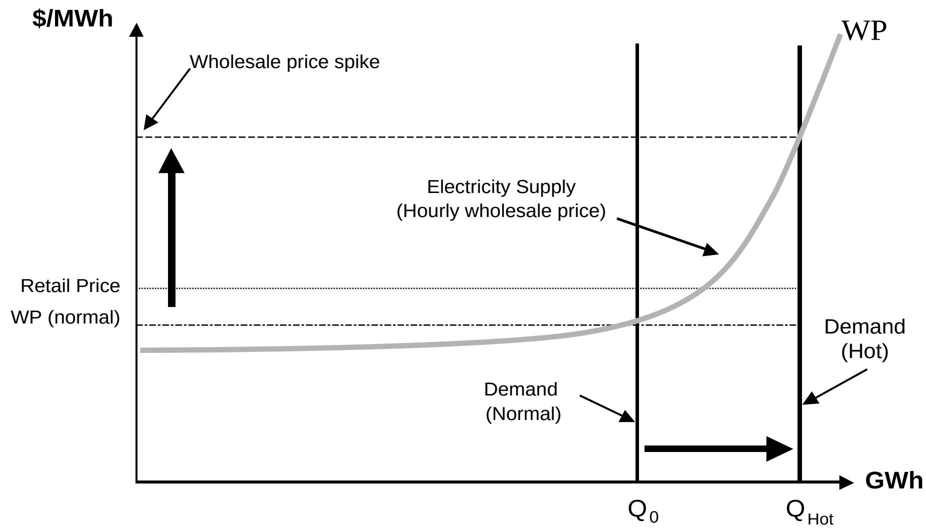
Electricity is essentially non-storable, which implies that it must be generated at the instant that it is demanded, thus requiring constant balancing of supply and demand by a system operator. Furthermore, customers' demand for electricity varies considerably over time, on an hourly, daily, and seasonal basis. As a result of these two factors, power systems tend to be characterized by a range of generation technologies that differ in terms of their capital and operating costs. These range from highly capital-intensive *base load* plants designed to run continuously at low operating costs, to *peaking* plants that are relatively inexpensive to install, and can start quickly to meet changing demand, but have high operating costs during the relatively few hours of the year that they are designed to run. Furthermore, to insure that capacity is available to meet demand during conditions of extreme load conditions or unexpected generator outages, a certain amount of reserve capacity is typically maintained.[4] These factors imply that the hourly marginal cost of electric energy, which reflects largely the operating cost of the highest cost generator that is dispatched to run, varies considerably across hours, days, and seasons.

This variability of electricity costs has been well understood by utility planning and operations staff. However, prior to the deregulation of wholesale power markets, these costs were largely internal to individual utilities and not visible in public markets. As wholesale power markets were opened up, generators offered blocks of power for various time periods at prices that generally reflected their operating costs. System operators matched supply bids to expected hourly loads, thus determining which generators to dispatch, and setting hourly market prices as the highest bids accepted. As a result, time-varying power generation costs became reflected in wholesale energy prices.

4 Rochlin (2002) points out that maintaining a reserve margin is the way that electric systems address the non-storable nature of electricity. That is, capacity is stored, rather than energy.

Figure 1 illustrates a typical electricity supply curve for a particular afternoon hour and the implied market price at two different load levels. At normal load levels, the wholesale price indicated by WP (Normal) might be around $50/MWh ($.05/kWh). On occasion, as demand in certain regions has soared due to extreme weather conditions and unexpected outages have constrained capacity, wholesale prices have spiked to levels exceeding $1,000/MWh ($1.00/kWh), reflecting not only the cost of the high-cost units that are the last ones called to run, but also the market's valuation of maintaining reliability and avoiding costly outages. Some of the high prices have also led to claims of market power on the part of some generators. That is, generators have been accused of "gaming" the system by withholding portions of their capacity in order to produce higher prices that they then receive on the remainder of their capacity.

Figure 1
Disconnected Electricity Markets

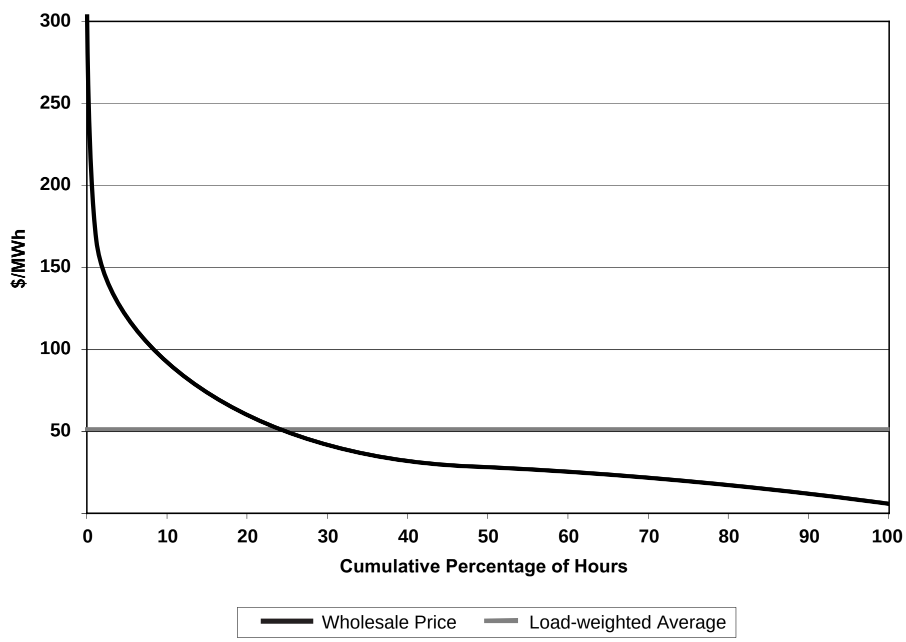


3.2 Traditional electricity pricing

Traditional utility rate design focused largely on recovery of allowed costs and on methods for allocating those costs fairly across various customer types. The economic efficiency of the resulting price structures, in the sense of establishing prices that reflect utilities' time-varying marginal costs, has typically been given a low priority. As a result, while *wholesale* costs vary hourly, *retail* prices for most customers differ seasonally at most (e.g., higher prices during the summer months than in non-summer months, due to higher costs driven largely by air conditioning loads). Only for large customers do prices typically differ by time-period during the day, in the form of time-of-use demand and/or energy charges. And only for a relatively few customers on real-time pricing (RTP) programs do prices vary hourly to match hourly wholesale costs.

Figure 2 illustrates the disconnection that can exist between varying wholesale energy costs and fixed retail prices. The curved line in the figure shows hourly wholesale prices in the PJM East region for the summer of 2000, arrayed from high to low. The flat line shows the load-weighted average of these prices, which turns out to be approximately $50/MWh (or $.05/kWh). This value represents a typical flat seasonal energy price that might be offered to small customers in this region. If charged in all hours, it would recover the same amount of revenue as if the actual variable prices were charged in each hour. Note, however, that in more than 70 percent of the hours of the summer, wholesale electricity costs were less than the average price (sometimes much less), while in more than 5 percent of the hours, electricity costs were more than twice as high as the average.

Figure 2
Distribution of Wholesale Energy Prices
(PJM East—Summer 2000)



Competitive markets for other commodities tend to produce prices that reflect production costs. The frequent wide differences between wholesale electricity costs and retail prices suggest extensive foregone opportunities for economic gain. That is, in an important but relatively small number of hours, the cost of producing electricity far exceeds customers' value of consuming it, as reflected in the price they willingly pay. Reductions in usage during these hours would save costs far in excess of customers' forgone value of power. However, it is also important to recognize that typical retail tariffs give customers no access to the relatively low-cost power that is available in the vast majority of hours. Increased usage during these periods would produce value to consumers that exceeds the cost of generating that power.

3.3 Market-based prices under retail competition

Under regulation, utilities are allowed to set retail rates to cover their expected energy costs, sometimes including fuel-adjustment factors that adjust rates periodically to reflect changing fuel prices. In an environment characterized by competitive wholesale and retail electricity markets, however, energy providers' perspective on cost recovery changes dramatically. The wholesale price curve in Figure 2 represents only one realization in a given year from an uncertain set of possible wholesale cost scenarios. Looking to the future, an energy supplier in a competitive market faces considerable risk due to uncertainty about future wholesale power costs. At the same time, most customers are likely to prefer the certainty of fixed, or guaranteed prices (e.g., a fixed price per kWh) for some time into the future.[5]

Load serving entities (LSEs)[6] will contract with those customers to provide power at a certain retail price, and then arrange to buy energy and ancillary services on the wholesale market to fulfill those contracts. LSEs face one dominant theme in deciding how to price their products to various customer types—how to manage the financial risk associated with uncertainty about future customer loads and wholesale power prices. That is, looking at a future time period, LSEs do not know exactly how much electricity each of their customers will consume, nor what the wholesale prices for that power will be at the time they will have to supply it.[7]

LSEs actually face three sources of risk from offering guaranteed prices—*wholesale price variability, load variability,* and *correlation* between wholesale prices and customer loads. First, they do not know what wholesale prices will be in the future, when they will have to purchase the power needed to meet their customers' demands.[8] Second, they do not know how much their customers will consume in any given time period in the future. Therefore, for example, they cannot enter forward contracts to meet all of their customers' demands; they will always have to purchase or sell back some power in spot markets. Finally, many customers' loads tend to be correlated with wholesale power prices. For example, residential and commercial customer usage tends to rise on hot summer days, which are the same time periods in which wholesale power prices tend to rise due to higher overall system loads. Thus, in the very hours in which wholesale prices are unusually high, many customers' loads will be unusually high as well.

---

5 An extreme example of potential risk is provided by the financial crisis faced by California's two largest utilities, which paid unexpectedly high wholesale power prices from May 2000 to early 2001, but had their retail prices fixed at low, regulated levels.

6 The list of potential LSEs may include the distribution components of regulated utilities, often referred to as utility distribution companies (UDC), that continue to serve customers in a service area, as well as competitive service providers who enter markets open to retail competition.

7 Utilities with provider of last resort (POLR) responsibilities also may not know how many such customers they may have to serve.

8 They can enter forward contracts to lock in prices for fixed amounts of power in the future. However, signing such contracts entails the risk that future spot prices may be lower than the forward price. This is the situation now faced in California as the prices for the long-term power contracts signed during the height of the crisis now appear much higher than current wholesale spot market prices.

Price-load correlation of this type makes the cost of serving certain customer types both higher and more uncertain than it is for others. The above three components of financial risk imply that LSEs will need to incorporate a risk premium into any guaranteed price offering.

Menus of retail price structures

Two particular types of price structures illustrate opposite extremes of the range of possible products that LSEs may offer to consumers. At one extreme of the range of price structures is a *guaranteed flat price*, such as a constant price in every hour of the period of the contract. In this case, customers may consume as much power as they wish at the guaranteed price. Under this arrangement, the energy supplier faces the entire wholesale price risk, and will need to include a *risk premium* in the price offered. Viewed alternatively from the customer perspective, this premium reflects the cost of insurance against volatile power prices.

At the other extreme are *spot prices*, in which the supplier offers to provide whatever amounts of electricity the customer wishes to consume at an hourly price that is tied directly to the wholesale price of power. This type of arrangement eliminates all risk to the supplier, who will be able to offer the product at little or no mark-up, needing only to cover his operating costs. The customer, however, bears all of the risk associated with uncertain wholesale prices.[9]

Between the two extremes of spot pricing and a single guaranteed price in all hours, one can imagine a wide range of possible price structures that have the effect of changing the allocation of risk associated with the factors described above. Two broad categories of such price structures are *guaranteed prices*, which are known in advance but may differ in certain time periods (e.g., flat, seasonal and time-of-use pricing), and variable, or *dynamic prices* that change on an hourly basis during at least some time periods to match changes in wholesale prices (e.g., real-time pricing and "critical" price TOU).[10] The guaranteed price structures will include a risk premium, as described above; dynamic pricing may not.

Dynamic pricing

The volatility of dynamic electricity prices suggests that many customers who are willing to forgo the risk premium of guaranteed prices in return for facing dynamic pricing will still want to engage in some type of price risk management. The standard approach to risk management that has developed in electricity markets is a combination of forward contracts and spot pricing known as contracts for differences. Hogan (2002) describes this type of price risk management in the context of transmission congestion pricing.

---

9 Financial risk management mechanisms, such as price caps and contracts for differences, may be employed to reduce that risk.

10 Eakin and Faruqui (2000) summarize the range of retail pricing strategies that are likely to evolve in a competitive retail market.

A straightforward example of dynamic pricing combined with risk management that has operated successfully in a regulated retail market environment is the *two-part RTP* service offered by Georgia Power Company and a few other utilities since the early 1990s. This tariff design consists of two parts. First, the customer pays hourly spot prices for its entire load (where the prices are typically based on hour-ahead or day-ahead wholesale prices, or forecast marginal costs). Second, a separate financial contract (a contract for differences) ensures that the customer pays no more than a given fixed price for a fixed amount of load (typically an historic baseline load). In the case of regulated utilities, the fixed price charged for the baseline load is determined by the customer's standard tariff.[11] This form of contracting should emerge naturally in a competitive market, except that the baseline load and standard tariff price will be replaced by the customer's choice of the amount of load to commit to a forward contract, and the forward price will be determined in financial forward markets, or agreed upon in bilateral contracts with generators.[12]

During periods of high hourly prices, customers facing dynamic prices, regardless of the extent to which they have managed risk through forward contracts, have an incentive to reduce consumption, thus reducing their bill. The beneficial effect that this demand response can have on wholesale market costs is illustrated in Figure 3. Rather than being fixed at one level, as in the vertical line in Figure 1, market demand it is now represented by the sloping demand curve shown in Figure 3. As higher wholesale costs are transmitted to those customers facing dynamic retail prices, their load reductions allow the market to clear at lower prices. The steeply sloping supply curve at low reserve margins has led to estimates that even modest load reductions (e.g., 2 percent of total load) can produce wholesale cost reductions that are ten-fold in magnitude (e.g., 20 percent), or more.

3.4 Traditional load management programs

The widely varying cost of electricity shown in Figure 2 has always given utilities an incentive to encourage some of their customers to reduce usage during periods in which costs were very high, especially in the relatively few hours in which reliability was endangered. The traditional mechanism has been load management programs such as direct load control and interruptible load programs.[13] Most programs have been designed as substitutes for peaking capacity; they are operated only occasionally during periods of unusually high loads or system emergencies. In most cases, customers have been offered a fixed discount or incentive payment in return for agreeing to have their load curtailed occasionally during emergency or high-cost conditions. The

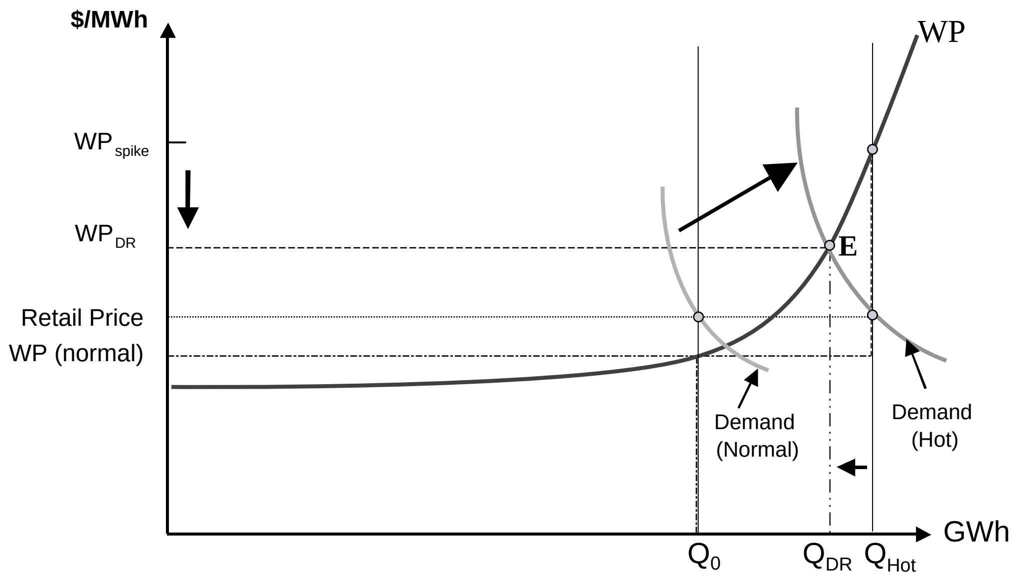
11 A technical description of the customer's bill under two-part RTP is provided in an appendix.

12 Georgia Power Company offers Price Protection contracts of this type to customers whose typical load exceeds its historical baseline load.

13 Direct load control programs typically involve devices that allow utilities to cycle or turn off residential appliances such as air conditioners and water heaters for certain fractions of each hour during designated time periods. Interruptible service programs typically involve large industrial customers selecting a firm power level to which they must reduce their load if the utility announces an interruptible event. They typically receive the rate discount on the difference between their firm power level and their typical load level.

amount of rate discount offered to interruptible customers has typically been tied to the value to the utility of the interruptible load, which is considered a partial substitute for peaking capacity. Thus, the rate discount has typically been tied to the cost of peaking capacity, adjusted for the extent to which interruptible load operates less frequently or reliably than a generator.

Figure 3
Effect of Demand Response on Wholesale Prices



4. CATEGORIES OF DEMAND RESPONSE PROGRAMS

For purposes of this paper we define *demand response* programs in general as mechanisms for communicating prices and willingness to pay between wholesale and retail power markets, with the immediate objective of achieving load changes, particularly at times of high wholesale prices. The ultimate objective of such mechanisms is to achieve the improved market performance discussed in Section 2.1. Various types of demand response programs have been proposed and/or offered. We adopt the classification scheme offered by Hirst and Kirby, who grouped demand response activities into three generic categories—*dynamic pricing, interruptible and voluntary load reductions*, and *customer provision of ancillary services*. We focus primarily on markets for energy, rather than ancillary services, and draw distinctions among three types of the second category of load reduction programs, as follows:

- The first category, *dynamic pricing*, provides customers with time-varying prices that reflect wholesale market costs. The most common example to date is provided by real-time pricing (RTP) programs for large commercial and industrial customers, in which hourly RTP prices reflect day-ahead or hour-ahead marginal costs. However, some examples of residential time-of-use rates with a "critical" price component that can be transmitted on short notice have been implemented, and are receiving renewed attention.

- *Interruptible and voluntary load reduction* programs involve customers offering to reduce their electricity usage during certain time periods in return for a financial payment. The load reductions may be voluntary or mandatory, depending on the type of payment, which may be a price discount offered in advance or a payment at the time of the reduction. Programs may be offered by incumbent utilities, LSEs, load or curtailment aggregators, or ISO/RTOs. Important differences between program types suggest subdividing this category into the following three types:

  – Traditional *load management* programs such as direct load control and interruptible/curtailable load programs can provide load relief to improve reliability during periods of low reserves or reduce cost at times of high wholesale market costs. Consumers have traditionally been paid in advance for their *participation* in such programs, through rate discounts or monthly bill credits, rather than for their *load reduction performance*. Load reductions are typically mandatory, with substantial penalties for non-performance. Some load management programs, particularly those that are rarely operated, have been criticized as serving as rate discounts in disguise. More market-based programs in the future are likely to combine pay for performance with payments for participation.[14]

  – In recent years, following price spikes in several regional wholesale power markets, some utilities have offered *energy buy-back programs*, in which customers agree to reduce usage in return for an incentive payment that is either established beforehand or tied to the wholesale market price. In some cases, these have been informal arrangements between utilities and large customers; others, such as the Energy Exchange programs operated in Oregon, consisted of formal arrangements that were activated whenever wholesale prices reached a certain level.

  – A number of *demand bidding* programs have been offered by the major ISOs, including PJM, ISO New England, the New York ISO, and the California ISO.[15] These allow retail customers or their aggregators, such as utilities or independent LSEs, to bid load decrements (i.e., load reductions compared to some baseline level) into day-ahead or hour-ahead wholesale energy markets. The load reductions are then scheduled and dispatched in a manner similar to the scheduling and dispatch of generators.

- In some cases, customers may bid load reductions into *ancillary services* markets, to provide different types of reserve services.

---

14 Efficient interruptible programs may offer a range of participation payments (the option payment) that vary inversely with market-based payments for performance (the exercise price). That is, to receive a large discount, a customer would have to curtail load at relatively low wholesale prices and receive low payments for the load reductions.

15 The term demand bidding is somewhat of a misnomer in that customers are actually bidding demand reductions.

Hirst and Kirby and Goldman et al. describe the features and recent performance of a number of programs in the first two categories. Hirst (2002b) discusses the role of demand response in markets for ancillary services.

Some forms of demand response would likely evolve naturally in the development of competitive wholesale and retail markets. However, as discussed in Section 5 below, continued of retail markets is likely to limit adoption of certain types of demand response mechanisms, such as dynamic retail pricing. Other barriers may exist in the form of market design rules, regulatory guidelines, or lack of infrastructure (e.g., advanced interval metering and communication systems).

Each of these categories of demand response mechanisms offers the potential to help connect the wholesale and retail energy markets. However, they differ in a number of key features, including the following:

- whether and how the amount of demand response is measured and/or validated,
- whether and how the amounts of retail demand response are taken into account in the wholesale market, and
- how the prices and/or incentive payments for demand response are set.

4.1 Measuring the amount of demand response

Measuring or validating the amount of demand response offered by customers or LSEs represents a key issue for the second category of *interruptible and voluntary load reduction* demand response programs. This is particularly the case for *demand bidding* programs in which payments are made for the amount of load reduction; that is, the amount by which load was reduced below a baseline level, which is designed to represent the customer's energy usage pattern in the absence of a request for load reduction. The fundamental problem is that load reductions cannot be measured directly. Only energy consumption can be metered; load reductions must be inferred by subtracting actual energy usage from a baseline level that is determined according to certain rules or algorithms. For example, a typical baseline load during a number of afternoon hours is calculated as an average of customers' usage during the same period on a number of previous days on which no demand response was requested or offered. A key issue in defining baseline loads is preventing gaming opportunities in which customers intentionally modify usage during the period in which the baseline load is calculated in such a way as to artificially increase the amount of load response for which they are compensated.[16]

16 An example of gaming is provided by the case in who customers' baseline load is determined by their load in the hour immediately preceding a demand response period. A customer who anticipates that the demand response will be called could increase his load substantially in the few hours prior to responding, thus establishing a baseline level above his normal usage, and receiving a demand response credit that does not reflect his typical usage pattern.

Measuring demand response is less of an issue for traditional utility *load management programs*, as payments are generally made in the form of rate discounts paid in advance, rather than pay for performance. Thus, measuring demand response represents a research and planning issue to determine appropriate rate discount levels. To the extent that these programs evolve toward a greater degree of pay for performance, measuring demand response will grow in importance.

Under *dynamic pricing*, customers are charged for what they consume, rather than for how much they reduce consumption. Since consumption may be readily metered, there is no need to measure individual customers' changes in usage from a baseline level. However, energy providers who offer dynamic pricing will have an incentive to develop an understanding of the *aggregate* demand response of its dynamic pricing customers so that they can provide accurate price-sensitive bids into the wholesale energy market. This will be especially important on days of high wholesale prices, when LSEs can avoid high-cost purchases to the extent that their dynamic pricing customers reduce their usage in response to price.

4.2 Accounting for demand response in wholesale markets

In order for retail demand response to actually affect wholesale prices, market mechanisms must be in place to reflect and anticipate the amount of demand response in the process of dispatching generators and setting prices. Different mechanisms may be needed for the different categories of demand response.

In a market in which some customers face dynamic retail prices, the natural way to account for demand response in the wholesale market is for LSEs to bid *price-responsive loads* into day-ahead and hour-ahead energy markets. For example, consider an LSE that expects to serve 5,000 MW of load during summer peak periods, 1,000 MW of which is represented by large customers facing day-ahead hourly prices. On a normal day, the LSE would schedule 5,000 MW each hour during the period, most of which would probably be covered by forward contracts. However, assume that the LSE knows that if prices rise to a level of $.20/kWh, then its RTP customers will reduce load by 100 MW, while at prices of $.50/kWh, they will reduce load by 250 MW. Thus, in scheduling loads into the day-ahead market, the LSE will offer to purchase 5,000 MW if prices remain low, 4,900 MW if prices reach $.20/kWh, and 4,750 MW if prices reach $.50/kWh. As the system operator matches load and generator bids in the process of determining the next day's unit commitments and prices, the price-responsive demand curves offered by LSEs would provide the mechanism for informing the market about the extent of demand response. This is essentially the approach followed by Georgia Power and the operations unit at Southern Company to anticipate the effect of RTP load response on day-ahead unit commitments; the result is lower day-ahead prices than if demand response were not accounted for.

Currently, few retail customers face dynamic prices, and, as noted by Hirst and Kirby, LSEs typically do not appear to bid price-responsive demands into the various energy markets. Instead, the recent focus has been on load reduction mechanisms by which large customers or LSEs are paid directly for load reductions in the day-ahead, hour-ahead, or real-time energy market. In this way, blocks of load reductions bid at various prices may compete directly with generators offering to supply power, with the result that demand response is accounted for in the wholesale market.

It has been suggested that demand bidding (i.e., load reductions that are bid or accepted at specific wholesale price levels) is the *only* mechanism that can assure that demand response is reflected in the wholesale market. In that view, customer response to dynamic prices is "passive" demand response that will not affect prices in the wholesale market. However, there is no reason that LSE's price-responsive demand bids will not convey the same information to the wholesale market and produce the same effect on wholesale prices as direct demand bidding, at least in the day-ahead and hour-ahead markets.

4.3 Pricing demand response

The amount of the financial payment that consumers should receive for load reductions under various demand response mechanisms has been the subject of considerable discussion. To provide guidance on this topic, in this section we examine the natural market incentives that exist for LSEs to pay their customers for load reductions under various possible retail contract forms. We first examine dynamic pricing and then turn to load reduction programs in which suppliers pay customers for demand response. Finally, we provide a numerical example.

Dynamic pricing

Customers facing *dynamic pricing* will pay a retail price that directly reflects wholesale market prices (e.g., RTP prices are typically set equal to marginal cost, or wholesale price, plus a small adder, or markup). During a period of high wholesale prices, they will face high retail prices that provide an incentive to reduce usage; for every kWh of consumption that they forgo using, their bill is reduced by the amount of the RTP price. These occasional *high* prices will be offset by relatively *low* prices that will hold in the majority of hours.

A dynamic pricing customer that has agreed to a simple *spot pricing* contract in which all usage is billed at hourly prices (e.g., day-ahead RTP prices) receives no specific financial payment from his supplier for load reductions; he simply has an incentive to adjust his usage according to the value of electricity compared to its price.

In contrast, the case of a dynamic pricing customer that combines spot pricing with a forward contract for price risk management provides a useful model for assessing incentive payments under alternative contract forms. This case may be illustrated by *two-part RTP*, in which customers pay standard tariff prices for a baseline load, and RTP prices for deviations from their baseline load. If wholesale prices, and thus RTP prices are high, and the customer reduces consumption below the CBL, then the customer's bill falls by the amount of the load reduction times the RTP price. This amount may actually appear on the customer's bill, and may thus be considered an implicit payment for the load reduction in the amount of the RTP price. It is important to note, however, that this payment is made after the customer has already purchased the baseline load amount at the tariff price. That is, the customer has purchased the baseline amount at a fixed price and then "sold" a portion of it back to the supplier by reducing usage below the CBL.[17]

Load reduction programs

For programs that involve payments for load reductions, the payment that customers receive for reducing load during a high-cost period depends on a number of factors, including the nature of their retail contract, the type of demand response program, and the rules of the program. In principle, market-based incentives exist for LSEs to make payments to customers for load reductions that effectively relieve the need to produce or purchase power at the high market cost. These are discussed below. Load reductions in these programs may be voluntary or mandatory (with penalties for non-compliance), depending on the nature of the program and contract.

A useful starting point for thinking about demand response payments is to refer again to the distribution of wholesale market costs shown in Figure 2. As described in Section 2, an energy provider that offers a guaranteed *fixed price* in all hours (or fixed TOU prices, as a similar curve exists for market prices during peak-period hours) must set that price to not only reflect a load-weighted average of the wholesale costs shown, but also include a risk premium to reflect uncertainty about the distribution of wholesale costs and the likelihood that loads will be high during high-cost hours. At that price, the supplier will be made whole, on an expected value basis, even though wholesale costs may occasionally reach high levels. However, a supplier may make itself better off by offering a demand response payment to the customer for load reductions during high-cost periods (see below). Under such a program, customer load reductions can be *voluntary* since the fixed price is already designed to cover costs during such periods.

---

17 As an example, with a CBL of 1,000 kWh, a tariff price of $.06/kWh, an RTP price of $.50, and a load reduction of 100 kWh below the CBL, the customer's RTP bill for one hour would equal 1,000 * $.06 -100 * $.50 = $60 - $50 = $10, which is $50 less than his bill for the CBL at the standard tariff price.

In contrast, consider an alternative contract that offers a *discounted fixed price* that applies in most, but not all, hours. In particular, during a limited number of hours the customer would either face market prices, or be required to reduce load to some predetermined level. The discounted price in this offering can be lower than in the previous example, because it does not have to protect the supplier against the occurrence of the very highest prices. In this case, however, the customer faces either *mandatory* load reductions (as in traditional interruptible load programs) or exposure to market prices during the high-cost hours in which the discount price does not hold.

Incentives for demand response payments—an example

The following example is designed to illustrate the incentives that may exist for utilities and LSEs to offer payments of different types to customers for load reductions during high-cost periods. The example considers such incentives under three alternative assumptions about the nature of the retail contract. First, assume that an energy supplier serves a customer whose typical afternoon load is 1,000 kWh in each hour, which is expected to rise to 1,100 kWh on a very hot summer day. On one such hot day, the day-ahead wholesale market price in the peak afternoon hours is expected to be $750/MWh (i.e., $.75/kWh).

Retail contracts. Consider the following three alternative retail contracts:

> *Hourly pricing*—the supplier simply passes the day-ahead hourly wholesale market prices (WP) to the customer (for simplicity, assume no markup).
>
> *Hourly pricing with risk management* (e.g., two-part RTP)—the customer pays a fixed price of $.08/kWh for a baseline level of peak-period usage (assumed to be 1,000 kWh per hour during the peak period). In addition, the customer pays (is paid) for any usage above (below) the baseline level at RTP prices.[18] For simplicity, assume that the RTP price equals the wholesale market price with no markup.
>
> *Fixed price*—the customer can consume whatever peak-period quantities it wishes, for a fixed price of $.10/kWh. Assume that the supplier has entered a wholesale forward contract to cover the customer's expected load of 1,000 kWh.

Load response. Assume further that in the two cases in which the customer faces dynamic hourly prices (i.e., *hourly pricing* and *two-part RTP*), the customer reduces usage from 1,100 kWh to 900 kWh during the period in which the price reaches $.75/kWh.

---

18 As noted above, this arrangement is equivalent to the customer paying hourly prices for all usage, and having a separate financial access charge, or contract for difference, that ensures that the customer pays no more than the agreed-upon fixed price for the baseline usage level.

Incentive for demand response payment. To examine the incentive that the supplier has in each case to pay the customer to reduce load, consider the effect on the supplier's net revenue in a single hour of the high wholesale price, and the effect of making a demand response payment:

Hourly pricing—the customer reduces usage in direct response to the high wholesale price, without any incentive payment from the supplier; the supplier's net revenue is unaffected since he simply passes on the wholesale price to the customer. The supplier has no incentive to pay the customer to reduce usage even more.

Hourly pricing with risk management (e.g., two-part RTP)—the RTP customer reduces usage in response to the high wholesale price. By the nature of the two-part RTP bill calculation, the customer's bill is reduced by the amount of his load reduction times the RTP price, or $150 (i.e., $.75 x 200). As under hourly pricing, this bill reduction represents a revenue reduction to the supplier. However, the supplier's cost is also reduced. He can either avoid the cost of buying the 200 kWh at the high wholesale price, or, if he has previously arranged a forward contract for that amount, like the fixed price supplier, he can sell it into the wholesale market. For the portion of the load reduction above the customer's CBL (i.e., 1,100–1,000), the effect is the same as for the hourly pricing customer: the customer avoids paying the high price and the supplier's net revenue is unaffected. However, by the nature of the two-part contract, the supplier has effectively paid the customer an incentive equal to the wholesale price for the portion of the load reduction *below* the customer's CBL (i.e., 1,000–900). That is, the customer has already paid the fixed retail contract price ($.08/kWh) for all consumption up to the amount of the CBL; thus, he essentially sells any usage below that level back to the supplier at the RTP price.[19]

Fixed price—the customer has no price incentive to reduce consumption below his normal level. The supplier is protected from the high wholesale price for the amount of the forward contract (1,000 kWh), but must buy the additional 100 kWh at the high market price, which he sells to the customer at the fixed price of $.10/kWh. Thus, he has an incentive to encourage the customer to reduce consumption. The question is how much he should be willing to pay. If he offers no payment, then he faces incremental costs of $75 for purchasing the 100 kWh beyond the forward contract (100 * $.75/kWh), and receives $10 for selling it to the customer (100 * $.10/kWh), for a net loss of $65. Thus, he will be willing to pay no more than $65 to encourage the customer to reduce consumption.

There are two ways to characterize this amount. One is that the supplier will be willing to pay no more than the difference between the wholesale cost and the fixed retail price for

19 Note that alternative RTP price designs are available in which the supplier pays less than the wholesale price for load reductions, thus earning a profit on such load reductions.

load reductions (i.e., \$.75 - \$.10/kWh, or \$.65/kWh). The other is that the supplier will require the customer to first buy an agreed-upon amount of load at the retail price, and then will be willing to pay no more than the wholesale cost for any reductions from that agreed upon amount. The agreed upon amount that the customer pays for at the fixed price represents the baseline load from which load reductions are measured in typical demand response program discussions.

To complete the fixed price example, the customer agrees to purchase 1,100 kWh at the fixed price of \$.10/kWh, the supplier then offers to pay the customer no more than the wholesale cost of \$.75/kWh for any load reductions below that amount. By the assumptions of the example, the customer reduces load to 900 kWh, and is paid an incentive of \$150 (i.e., \$.75 x 200). The supplier avoids the cost of purchasing the 100 kWh in excess of his forward contract, and has now freed up 100 kWh from that contract that he can now sell into the wholesale market. In summary, under a fixed price contract, the supplier has an incentive to pay the customer for load reductions below a baseline level for which the customer has paid the fixed retail price, in an amount no more than the wholesale market price. This is the case even if the supplier is covered by forward contracts, because he faces an opportunity cost of *not* offering to do so. Finally, since suppliers will need an incentive to do better than breaking even, they will have an incentive to pay customers something less than the wholesale price, thus allowing them to earn a margin on sales back into the market. In this way, both the consumer and supplier may be made better off.[20]

Conclusions. Under both versions of *hourly pricing*, in which the price to the customer reflects the high wholesale cost, the customer has a direct incentive to reduce load and the supplier may have no incentive to offer any additional incentive for its customer to reduce load further. Under *two-part RTP*, the supplier has already effectively sold the customer his baseline load at the fixed price, and the balancing contract implies that the supplier will pay the customer the RTP price whenever he reduces load below the baseline level. Finally, under a typical *fixed price* contract, the supplier may not see an obvious incentive to pay customers for reducing load, particularly if he is fully covered by owned generation or a forward contract at a price below the temporarily high wholesale price. However, an opportunity exists for gains to both customers and supplier from load reductions during conditions of high wholesale market costs. The key is to agree on a way to calculate a baseline level of consumption for which the customer pays the fixed retail price and a demand response payment no larger than the wholesale cost of power. The supplier faces an *opportunity cost* of not offering such payments.

20 This is analogous to the case of a traditional utility that has adequate generation during a period of high wholesale market prices, and offers to "buy back" power from certain large customers at a fraction of the market price, thus sharing with the customer the profit from selling the freed-up power into the wholesale market.

The lesson from these examples is that competitive power markets can provide incentives for suppliers to pay customers to reduce load during periods of high wholesale prices. It is important to note that all of the payments in the example are market based; they are tied directly to the wholesale cost of power, rather than set arbitrarily or administratively, and do not exceed the wholesale market price (after the customer has paid the retail price for the baseline load from which load reductions are calculated). This profit incentive may be most apparent in a competitive retail market. Regulatory rules and ratemaking may present barriers to regulated utilities having the same incentive to take such actions. Furthermore, revenue recovery issues exist for distribution utilities that price distribution services largely through volumetric charges; they may face revenue losses on the wires component of an unbundled tariff due to the load reductions that occur as a result of demand response payments. That is, the fixed retail price in the example refers to the price for only the unbundled energy portion of the customer's bill.

Demand response at low market prices

One topic that receives little attention in most discussions of demand response is the benefits that can be generated by customer demand response during periods of *low* wholesale market prices (i.e., *load increases*). That is, most discussion focuses only on the benefits to consumers to be derived from load reductions at times of high wholesale market prices. Reference to Figure 2, however, which shows the frequent occurrence of low market prices, serves to indicate the potential benefits to be derived from customers increasing usage at times of low market prices, and shifting load from high-cost to low-cost periods. Note that these benefits will likely only be achieved through dynamic pricing, as demand bidding and load management programs are typically designed to operate only during infrequent periods of high wholesale prices.

5. HISTORICAL EVIDENCE ON PRICE-RESPONSIVE LOADS

The question of how much load response can be expected from demand response will depend on what types of mechanisms are implemented and customers' willingness to change usage patterns in response to price incentives. Considerable information has been compiled about the performance of several utility and ISO-sponsored demand response programs during the summers of 2000 and 2001. In addition, a considerable literature has developed over the past 25 years documenting customer price responsiveness to time-varying prices. The following sections offer illustrative examples.

5.1 Recent demand response program results

Hirst and Kirby reported on a number of demand response programs in the summers of 1999 and 2000, while Goldman et al. summarize a number of program results for the summer of 2001. The experience from the most recent summers is limited by the relatively few periods of system emergencies or high wholesale prices in the Eastern United States. Some examples of the magnitude of load response reported are the following:

- Since 1991, PJM has sponsored an Active Load Management program which is operated in the form of direct load control and interruptible load programs by the distribution utilities, which receive installed-capability credits for load reductions. The program achieved load reductions in the range of 1,700 MW, or 3.5 percent of PJM's summer peak demand, in 1999 but was not implemented in 2000.

- PJM, NY ISO and ISO New England have all operated voluntary load reduction programs in which LSEs receive payments for load reductions at prices tied to market prices or set administratively (e.g., a minimum of $500/MWh or the locational marginal price of energy). Some are triggered by reliability conditions, while others are based on market price levels. Some of the programs report load reduction capabilities, on the order of 100 to 250 MW. The NY ISO Emergency Demand Response Program, was reported to have provided 425 MW of load response on four occasions in 2001.[21]

- Cinergy has reported signing up a large portion of its large commercial and industrial customers in a range of load reduction programs under the umbrella *PowerShare Pricing Program*, with an estimated curtailable load ranging from 440 to 600 MW. However, the program was not invoked in 2000 or 2001 due to low market prices.

5.2 Evidence of customer price responsiveness

Kahn (2002) recently pointed to the historical tendency in the electric power industry to underestimate and even doubt the extent of customer response to price changes. However, dozens of studies over the years have documented reasonably consistent patterns of customer price responsiveness, particularly to time-varying prices such as TOU and RTP (see EPRI [2001]). Most relevant to discussions of demand response during infrequent periods of high prices are recent studies of customer price responsiveness in cases in which prices change with relatively short notice. A few such studies are summarized below.

Real-time pricing

A number of utilities in the U.S. have experimented with two-part RTP along the lines of Georgia Power, including Niagara Mohawk Power Corporation, Duke Power Company, Cinergy, Utilicorp, Xcel (Public Service Company of Colorado), and Kansas City Power and Light. The largest and most extensively analyzed RTP program in the U.S. is that of Georgia Power Company. It has grown from a pilot program of approximately 30 customers in 1993, to a permanent program

21 See Neenan (2002).

with some 1,700 customers and 5,000 MW of load in 2001. Georgia Power has conducted periodic analyses of RTP customer price responsiveness in the course of developing and maintaining an RTP load response forecasting model that can be used to anticipate RTP load response at varying price levels for both system planning and operations functions.

Braithwait and O'Sheasy (2001) document recent estimates of Georgia Power RTP customer price responsiveness, by customer type and price level. Table 1 shows estimated hourly own-price elasticities (i.e., the ratio of the percentage change in hourly usage to the percentage change in hourly price) for seven customer segments, at a range of RTP prices, based on an analysis of data for 1999, during which several extremely high-price hours were observed.

Table 1
Georgia Power Company RTP Price Elasticities
(by customer type, at alternative price levels)

	Elasticities by Customer Segment						
Price ($/KWh)	**HA**	**NSEI**	**NSEC**	**SEI**	**SEC**	**OSI**	**OSC**
0.15	-0.180	-0.019	-0.014	-0.066	-0.039	-0.077	-0.073
0.25	-0.206	-0.029	-0.027	-0.084	-0.052	-0.073	-0.089
0.50	-0.241	-0.044	-0.045	-0.109	-0.069	-0.067	-0.110
1.00	-0.276	-0.058	-0.062	-0.134	-0.087	-0.061	-0.132

The customer segments were defined as follows:

HA: Very large customers facing hour-ahead RTP prices

OSI: Industrial customers with on-site generation

OSC: Commercial customers with on-site generation

SEI: Industrial customers previously on a supplemental energy (SE) interruptible tariff

SEC: Commercial customers previously on a supplemental energy (SE) interruptible tariff

NSEI: All other industrial customers

NSEC: All other commercial customers.

The hour-ahead, on-site generation, and formerly interruptible industrial segments were found to be the most price-responsive. The two Non-SE segments were only slightly responsive to prices below $.25/ kWh, but increased their responsiveness substantially at the higher prices. Overall, Georgia Power reports an RTP load response capability of at least 750 MW at the occasionally high prices observed in 1999.

Duke Power Company, which operates a similar though smaller RTP program, reports load reductions of about 200 MW at high prices from approximately 50 very large customers.

Residential "critical price" TOU

Over the last decade, a few utilities have tested a unique type of residential TOU rate that combines a standard peak, off-peak, and shoulder-period TOU rate with the ability to make the peak price dynamic to reflect occasional high wholesale market prices. The dynamic pricing is implemented through advanced communication and control technologies. The programs provide each customer with an interactive communication system which allows the utility to send out a "critical" price signal to participating customers during infrequent periods of high-cost supply conditions (critical price levels have typically ranged from approximately $.25 to $.50/kWh, and many programs limit the number of critical price hours to no more than 2 percent of all hours). The systems also allow customers to pre-schedule their thermostat settings and certain other circuits in response to both the standard TOU price periods and the receipt of a critical price signal.[22]

American Electric Power (1992) reported peak demand reductions ranging from 2 to 3 kW per customer at high prices, and 3.5 to 6 kW at "critical" prices.

Braithwait (2000) analyzed participant and control group load data for a pilot program at GPU Energy and confirmed that customers modified their usage patterns substantially in response to the TOU and critical prices. Customers *reduced* consumption during peak periods and some shoulder periods, and *increased* consumption during off-peak and some shoulder periods. Summer peak-period load reductions averaged about .5 kW, or 25 percent of control group loads, while response during critical price periods ranged from .6 to 1.24 kW. Estimated elasticities of substitution were larger than in most previous studies of traditional TOU programs, indicating strong customer price responsiveness. Specifically, an elasticity of substitution of 0.31 was obtained from a constant elasticity of substitution (CES) demand model, which contrasts with estimates from traditional TOU programs that average around 0.15.[23]

International evidence of price responsiveness

Aubin et al. (1995) reported finding strong price responsiveness and substantial net economic benefits in an experimental residential real-time pricing program offered by EdF in France that was made possible by recent reductions in metering costs. Peak and off-peak prices were established for three *day-types* (e.g., low, moderate, and high-cost). Customers were notified of the next day's day-type by eight p.m. (through the meter). The utility allocated a limited

---

22 Gulf Power Company currently offers an experimental TOU program of this type in Florida.

23 An elasticity of substitution characterizes the degree of load shifting from one time period to another (e.g., from the peak to off-peak period). Technically, it represents the negative of the ratio of the percentage change in quantity to the percentage change in price between two time periods.

number of high (22) and moderate (43) days throughout the year. According to the EdF web site, the program continues to operate under the Tempo name. This type of rate design allows lower peak period prices on the low and moderate-cost day-types than under typical TOU rates, and provides strong incentives for customers to reduce load during the relatively few high-cost peak periods.

6. INCORPORATING DEMAND RESPONSE IN MARKET DESIGN

Demand response has been identified as a critical link in connecting retail and wholesale electricity markets. The absence of retail demand response has been indicted as the chief culprit behind the price spikes, shortages, and charges of market manipulation that have plagued several U.S. electricity markets the past few years. The call to provide demand response a role on equal footing with supply-side resources has been clearly heard and widely accepted as imperative. The challenge and controversy that remains is how to incorporate demand response into market design to achieve the most efficient and effective market performance.

This section tackles the question of the appropriate role of demand response in market design. To address the issue, we begin by describing the market structure and the key market participants. We then provide a vision of how demand response would manifest itself in a competitive market structure. Using the competitive market analysis as a benchmark, we then address the challenges of incorporating demand response into retail electricity markets that remain largely regulated. Finally, we return to the fundamental questions, challenges, and basic design issues that were identified in Section 2 and offer answers and recommendations.

6.1 Market structure and participants

As the electric power industry has proceeded through the restructuring process this past decade, electricity markets have been (somewhat artificially) separated into distinct wholesale and retail markets. The wholesale markets are being organized at the RTO/ISO level. The wholesale market authority (e.g., the RTO/ISO) has the general responsibility for providing a non-discriminatory market environment that promotes economic efficiency, maintains power system reliability, lowers the cost of delivered energy, and increases choices offered to wholesale market participants. In addition, the market authority has the duty to monitor market performance and mitigate the exercise of market power.

Market participants

The primary participants in the wholesale market are owners of supply-side resources (e.g., generators), and load serving entities (e.g., investor-owned utilities, municipalities, and other retail aggregators), marketers, and some large direct-serve commercial and industrial customers.

Companies that transport power (transmission and distribution companies) also participate in the wholesale markets. Wholesale market participants range from the traditional vertically integrated utility to the independent single generator and the small merchant retailer. Proposed wholesale market design rules would require an independent entity to serve as the market authority, called the “transmission provider” (FERC 2002a, p. 5).

Regulators also participate in wholesale markets. FERC has primary regulatory jurisdiction over wholesale markets. The RTOs and ISOs receive their market authority from the FERC. State regulators also have some regulatory authority in the wholesale markets. In states like California, New York, and Texas, the state regulatory role may be relatively larger (*vis-à-vis* FERC) because the wholesale market is intrastate (e.g., CAISO, NYISO, and ERCOT).[24] Likewise, the state regulatory role in wholesale markets remains strong in states where the electric industry remains dominated by vertically-integrated utilities that are not organized into ISOs or RTOs. As discussed below, the presence of multiple regulators makes the incorporation of demand response into wholesale markets a more complex challenge.

Wholesale markets

Wholesale markets consist of energy, transmission congestion, and ancillary services markets. Wholesale energy markets are typically made up of several different time-differentiated markets, including long-term block forward markets, and day-ahead, hour-ahead, and real-time energy markets with hourly-differentiated prices. Ancillary services markets (especially operating reserves) are differentiated by response time. The wholesale energy and operating reserves markets are inter-related because resources can be allocated to more than one of these markets, depending on their flexibility in adjusting output or load levels. Thus, market design and pricing in these markets are not independent.

The LSEs and the direct-serve customers, being the demand-side of the wholesale markets, are the direct sources of demand response in the wholesale markets. End-use retail customers can indirectly provide demand-side resources to wholesale markets through their LSE, or possibly some other aggregator.[25] Ancillary services are currently provided almost exclusively by generators. However, demand-side resources can in principle provide some ancillary services.[26] There are several distinct reserve services markets, differentiated by the time required to get the service on-line (10-minute, 30-minute, etc.). The ability of demand resources to participate in these markets increases with the response time.

---

24 ERCOT differs substantially from California and New York given the lack of significant connections for imports and exports.

25 As discussed below, having retail load delivered to the wholesale market by an entity other than its load serving entity can create revenue recovery problems in regulated retail markets.

26 Hirst (2002b) provides recommendations on how to modify current operating rules to allow greater opportunity for demand response to contribute to ancillary services, as well as energy markets.

Retail markets

Retail electricity markets remain under the jurisdiction of state regulators. While some states continue to move toward retail competition (Texas is currently on the front line), some have retreated back to a regulated retail environment (most notably, California). The retail electricity market is typically a single market for delivered energy. The primary participants are the LSEs and the end-use customers purchasing from the LSEs. Distribution companies also participate in the retail market. Some LSEs also own the distribution wires, while other LSEs are pure merchant firms. Secondary participants in the retail markets are distributed generation resources which may provide supply resources into the wholesale market by accessing the distribution network or who may indirectly provide retail demand resources by replacing power purchased by an end-user through an LSE.

Retail electricity markets traditionally have been characterized by consumers making usage decisions in real time, based on regulated prices established in advance and typically not differentiated by time period. Consequently, there has been little incentive for retail load to respond to wholesale costs. The lack of price response in retail markets is not a result of customers being price insensitive—they aren't. Instead, price response in retail markets is absent largely because retail prices generally do not vary; therefore, otherwise price sensitive customers have no incentive to respond. The demand response challenge with respect to retail markets is not so much how to make customers more price responsive, but instead how to make retail pricing structures more dynamic.

6.2 Demand response in competitive markets

One natural way to assess alternative demand response approaches is to first consider what form demand response would likely take in a fully competitive market. That is, what mechanisms might logically evolve in power markets characterized by competition in both the supply of and demand for energy, where the market design would also include a regional system operator that manages the markets and schedules generators according to bids for demand and supply.

Analyzing how demand response would work in a competitive market environment is not meant to suggest that competitive retail electricity markets are either imminent or inevitable. Neither is the analysis meant to suggest necessarily how demand response "should" be incorporated in actual electricity market design. Instead, the analysis of a competitive market structure is meant to serve as a benchmark from which the impacts of real-world market barriers and regulations can be discerned, and to provide insights into implications for market-based demand response mechanisms.

Competitive markets for any commodity incorporate demand response naturally. In competitive markets, market participants face price risk exposure and have incentives to respond to changing market conditions. Information on changing market conditions is conveyed by market prices. As market conditions tighten, prices rise and purchasers respond by buying less, other things the same. Based on individual tolerances for risk exposure, market participants in a competitive market use hedging instruments (e.g., forward contracts and price options) to manage their price risk.

In competitive *wholesale* electricity markets, typical price hedges consist of forward contracts and price options that apply to specific amounts of load. That is, wholesale forward contracts have the traditional elements of a pre-specified price for a specific amount of load to be delivered and paid on a specific date (e.g., a block of 1,000 kWh each hour between 8 a.m. and 8 p.m. on weekdays in July). Price options, such as price caps and price floors, will provide price protection on a specified amount of load. Wholesale hedging activities affect a market participant's risk exposure, but do not affect the perception of costs and value *at the margin*. Consequently, in competitive electricity markets, virtually all of wholesale demand is effectively subject to dynamic pricing.

At the *retail* level, however, some (probably most) customers are likely to desire price insurance for *all* of their consumption. These customers will not respond to changes in wholesale prices because their retail prices will not reflect those changes. However, these customers will pay a market-based (i.e., actuarially-fair) premium for the price insurance. Some retail customers will decide to accept some wholesale market risk (i.e., self-insure) to avoid the insurance premium. As more retail customers self-insure, facing dynamic prices and providing demand response, wholesale market price volatility will decrease, as will the retail price insurance premium, as a result. The decrease in the price insurance premium makes the guaranteed price product look more attractive and decreases the incentive to self-insure. Thus, in competitive markets, the amount of retail load resources providing demand response will be a market-determined equilibrium outcome rather than a specific amount or policy mandate.[27]

Competitive wholesale markets

Essential to competitive markets is the development of spot markets. Given the idiosyncrasies of electricity (e.g., non-storability, need for continual system balance, etc.), at least three types of "spot markets" exist for wholesale electricity. These are the day-ahead hourly market, the hour-ahead market, and the real-time market. These spot markets provide the foundation for longer-term wholesale forward contracts. Forward contracts provide the primary mechanism for

27 Glyer (2000) has illustrated the incentive that all interval metered customers will have to reduce load during high-cost hours, regardless of whether they face fixed or dynamic prices. The incentive is indirect, and comes with a time lag. The source of the incentive is that suppliers will have an incentive to offer customized flat or TOU prices whose level depends on each customer's actual historical usage pattern. Thus, customers will have an incentive to reduce usage during high-cost time periods to reduce their expected price in future periods.

the wholesale market participants to hedge against price risk. In addition, a robust forward market reveals market-based price volatility information.[28]

Transparent market volatility measures facilitate the development of additional price risk management options (e.g., price caps, floors, and swaps). In wholesale markets, forward contracts and price options apply to pre-specified quantities of load. Forward contracts and price options allow market participants to individually tailor their exposure to wholesale price risk. However, the use of these hedging instruments does not change the incentives of the wholesale purchasers to provide demand resources back into the wholesale market when wholesale prices get high. Wholesale purchasers will compare the *marginal* value of usage to the wholesale price. If the wholesale price exceeds the marginal value of usage, the wholesale purchaser would increase profit (net benefit) by reducing usage in the case of a direct-serve industrial user or, in the case of a load-serving entity, by paying the end-use customer to curtail.

The interaction of competitive wholesale energy and reserve markets

In competitive wholesale markets, competitive prices for ancillary services and for energy services are closely related. This is because these markets represent production alternatives for supply-side resources. With competition and freedom to allocate generation capabilities across these markets, the prices in the markets should equilibrate.

The competitive pricing structure for reserve services will likely consist of two components—an availability price and an exercise price for actual performance. Availability will involve some advance commitment (such as day-ahead), and therefore the competitive availability price will be closely related to the corresponding energy price (minus the fuel cost) in the corresponding market. This is because the opportunity cost of committing the resource to the reserve services market is the forgone profit in the energy market. The exercise price for reserve services will be most closely related to the hour-ahead and real-time energy prices.

Demand resources can participate in many of the ancillary services markets, with required response time being the key determinant of the demand-side potential.[29] However, demand response participation in the energy markets may indirectly improve performance in the ancillary services markets (even the instantaneous response market for regulation services). This is because demand participation in energy markets puts downward pressure on energy prices and "pushes out" some supply-side resources that are better adapted to rapid response in the reserve services markets.

---

28 In robust forward markets, day-to-day price changes for the same forward contract are by definition unanticipated and therefore reflect the pure volatility of prices. In contrast, extracting volatility measures from forecasts of spot prices requires econometric de-trending to isolate the random components which contain the volatility information.

29 See Hirst (2002a) for further discussion of the demand-side potential in ancillary service markets.

Furthermore, in a competitive market structure, the distinction between reliability and economic needs will tend to blur, leading to a further blending of the two reserve services and energy services markets. As demand resources contribute to improve market performance, greater price and load stability are likely results. Consequently, the desired or needed reserve requirement may actually decrease. If so, then demand response in energy markets may indirectly reduce demand pressure in the reserve services markets.

Competitive retail markets

The primary market participants in a competitive retail energy market are end-use customers on the demand side and LSEs on the supply side. Competitive retail markets are characterized by customers having choices among several LSEs, and across a variety of risk-differentiated products. Competition among LSEs for customers will lead to the development of those differentiated pricing products, including products that in essence provide demand response opportunities in the wholesale markets (see Section 3.3). The responsiveness of LSEs to the desires of retail customers with competitive alternatives will diminish the role that exists for aggregators in non-competitive markets to fulfill retail customer needs left wanting by unresponsive LSEs. Likewise, as competition eliminates non-market-based rates and revenue recovery mechanisms, some artificial roles for non-LSE aggregators will disappear.

Retail products in competitive markets involve combinations of the electricity commodity (electrons) and some degree of price insurance. At one extreme of the retail product spectrum is unhedged spot pricing and at the other extreme is the fully insured guaranteed price product, or perhaps even a fixed-bill product.[30] In the middle of the retail product spectrum will be partially-hedged products such as two-part real time pricing.[31] Competitive markets will have market-based pricing of all the retail products, whether fixed bill, guaranteed price, or two-part RTP of un-hedged spot.

The importance of forward contracting in retail markets

As described in Section 3.3, the inherent volatility of wholesale electricity prices suggests that most market participants, including consumers on the demand side and suppliers on the supply side, will want to manage their price risk. The standard approach to price risk management in competitive commodity contracts is forward contracting. That is, parties agree to buy and sell fixed quantities of electricity at a given price at some time in the future. Forward markets provide a mechanism for forward prices to be established based on participants' expectations of future prices. Since consumers can never be certain that their usage will exactly

---

30 A fixed bill product consists of a simple monthly charge that does not vary with usage.

31 Two-part real-time pricing consists of a forward contract with deviations in usage balanced at spot prices.

match the contract amount, forward contracts will be associated with balancing contracts that establish prices (e.g., hourly dynamic prices tied to the wholesale market price) for deviations in usage from the forward contract amount.

Section 4.3 illustrated that during periods of high wholesale prices this combination of contracts operates similarly to a load reduction type of demand response program in which customers are paid incentives to reduce consumption. That is, customers who have entered forward contracts to purchase a baseline level of consumption at a fixed price are effectively paid at wholesale spot market prices for any load reductions below their baseline, forward contract load level. There is a key difference between this type of contract arrangement in a competitive market, and most load reduction demand response programs—the baseline load level is established in advance and is chosen voluntarily by the customer. This is in contrast to the difficult, and sometimes contentious measurement issues involved in many demand response programs.

Demand response mechanisms in competitive markets

We now return to the categories of demand response programs that were presented in Section 4. By drawing upon knowledge of the peculiarities of electricity and observations of competitive commodity markets, we speculate how demand response programs would evolve in a competitive electricity market structure. In each case the forward contract plays a central role in the competitive version of the demand response program.

Dynamic pricing, particularly hourly real-time pricing, is the pricing structure most consistent with how other commodities are bought and sold in competitive markets. Competitive wholesale electricity markets already provide wholesale customers with block-forward contracts, as well as real-time, hour-ahead, and day-ahead hourly prices. At a competitive retail level, the largest and most electricity-intensive industrial customers will likely purchase electricity in a similar manner with forward contracts for their expected baseline needs with deviations settled in the day-ahead or hour-ahead market.[32] This dynamic pricing structure is similar to how industrial customers purchase their other essential inputs.[33] In this pricing arrangement, the customer essentially purchases all electricity at the "real-time" price, and uses the forward contracts as pure financial hedges to reduce risk. The customer's LSE takes into account the expected load changes of all of its dynamic pricing customers in bidding price responsive loads into the wholesale market.

Interruptible and voluntary load reduction programs may co-exist with dynamic pricing in competitive electricity markets, but will be structured quite differently from their current forms. In fact, the competitive structure of these programs makes them very similar to dynamic pricing.

32 The size of the forward contract will depend upon the customer's overall risk management needs. Most "real-time" balancing of actual usage will occur in the day-ahead market, with some of the most flexible customer choosing to settle in the hour-ahead market.

33 For example, how metal product fabricators purchase metals, how cereal producers purchase grain, and how meat processors purchase meat.

In particular, with competitive market-based pricing, the pricing structure will likely involve more pay-for-performance (i.e., payments for the amount of load reductions during periods of interruption, including reductions greater than the contracted amount) and less up-front compensation for participation (i.e., rate discounts in advance for the right to interrupt under certain conditions). In addition, the issue of establishing a baseline from which to measure demand response will largely be addressed by the use of forward contracts. For *interruptible* programs where the LSE buys call options for load reductions from retail customers, the customers wanting to participate will choose the amount of load they want to be eligible for interruptible programs by purchasing forward contracts, which can be transferred to the LSE if an interruption is called. Interruptible contracts with a select group of large customers may be useful to the LSE that wants to participate in the reserves services markets. Similarly, *energy buy-back* programs, where the customer has the opportunity but not the obligation to reduce usage for a market-based payment, will likely require forward contracts to establish the baseline usage.[34]

Demand bidding programs sponsored by ISOs/RTOs at the wholesale market level may still exist in competitive electricity markets. LSEs will bid load reductions into the wholesale market based upon the expected load responses from their customers. These load responses may come in part from customers' direct responses to the incentives offered by the ISO, and passed through by the LSE, with some sharing of the incentive, as is the current practice. However, they may also come from the LSE's forecasts of customer load changes from its dynamic pricing, interruptible load, and energy buy-back customers. In competitive markets, however, there will be no obvious role for other aggregators to go directly to retail customers. Those retail customers on dynamic pricing and interruptible programs will already be participating, indirectly, via their LSE.

Competitive *ancillary services markets* will have demand-side participation via the LSE bidding. The LSE will probably bid loads into the reserve services markets based on special contracts with select retail customers that are particularly well-suited to rapid response to a curtailment call. However, the amount of demand-side load bid into reserve services markets may be relatively low, as some additional supply-side resources may be shifting to reserve services markets from energy markets as energy market demand becomes more price responsive.

6.3 Demand response with non-competitive retail markets

The previous section concluded that a combination of competitive wholesale and retail markets would provide natural incentives to encourage demand response, in the form of dynamic pricing and some types of load reduction programs. Given the choice, some customers would be willing to face dynamic retail prices that vary with wholesale prices. During periods of high wholesale prices, these customers would have a financial incentive to reduce load, and their

34 Energy buy-back programs are similar to the practice in the U.S. airline industry of "voluntary bumping" where in an overbooked situation the airline pays a ticketed customer to take a later flight. In this case the baseline (a seat) is a well-defined property right established by a forward contract (the passenger ticket).

LSE would have an incentive to bid a price-responsive load into the wholesale market. That is, a demand response mechanism ought to arise naturally through the incentives provided by competitive markets. Currently, however, few retail electricity markets may be classified as fully competitive. Most customers continue to be served by utilities that remain regulated by state public utility commissions. Even in states with retail access, in which customers have a choice of supplier, regulated *standard offer* and *provider of last resort* rates provide an important benchmark to which competitive prices must be compared.

This section examines some of the challenges and barriers to demand response mechanisms in a transition period with varying degrees of retail market regulation, the incentives that exist for different types of demand response, and some market-based design features of demand response mechanisms, using examples from the competitive framework as a benchmark.

Challenges and barriers to demand response

In the current transition period, a number of challenges and barriers limit the extent of demand response mechanisms that contribute to the operation of wholesale markets. We group these barriers into three categories—regulation, technology, and measurement and other implementation issues.

Regulation

The primary reason for the current lack of price responsive load is presumably the same one that has held historically—the lack of incentives provided by traditional cost of service-based rate making. That is, utilities are allowed to collect revenues designed to cover their expected costs, including those in both the many low-cost hours and the occasional high-cost hours that were illustrated in Figure 2. Some also have fuel adjustment mechanisms that automatically adjust rates over time to reflect unexpected changes in fuel prices. Thus, most utilities have little incentive, or are not permitted, to offer innovative pricing strategies that give customers choices about the amount of price risk they are willing to face. At the same time, regulators see a political need to protect customers against volatile prices. Thus, customers see stable prices in return for paying a rate premium on average to cover the capacity reserves needed to maintain reliability in the absence of price responsive load.[35]

Even in states with retail access, the typical design of the regulated *standard offer* and *provider of last resort* rates limit incentives for providers to offer, and customers to choose, dynamic pricing.[36] Customers see today's standard offer prices as safe havens to which they can always

---

35 Note that many, if not most, large commercial and industrial customers face time-varying prices in the form of time-of-use energy and/or demand charges. These prices typically vary in approximately the same pattern as wholesale prices; however, they do not increase dynamically to reflect unusually high wholesale prices during periods of system constraints.

36 Hirst (2002b) discusses how rate discounts guaranteed in typical utility standard-offer prices undermine incentives for dynamic pricing and retail competition.

return if market prices turn out unfavorable. This discourages them from committing to alternative combinations of dynamic pricing and risk management contracts that might otherwise be attractive when compared to a market-based standard offer.

A further feature of standard cost-of-service tariffs that provides a disincentive to dynamic pricing and other demand response mechanisms is the common practice of recovering most revenue through volumetric energy charges. That is, the cost to utilities of providing some services, such as energy, varies directly with usage. When energy costs are high, a customer's load reduction can reduce the utility's cost accordingly. However, the cost of providing other services, such as the distribution wires service, is largely fixed. Attempting to recover those costs through volumetric pricing of distribution services can lead to under recovery if customers respond to high *energy* prices by reducing consumption. One method of recovering fixed distribution wires costs, which vary by customer, is through a monthly access charge. However, practical problems in determining customer-specific fixed charges suggest the use of demand charges to recover distribution wires revenue for those customers with demand meters. Kirsch and Hemphill (2000) have reviewed approaches for redesigning distribution tariffs in the face of competitive market forces.

Technology

The primary technological barrier to all types of demand response mechanisms is metering and communications. Charging or compensating customers for usage changes during specific hourly time periods requires hourly metering. Informing customers of dynamic prices or bid offers requires some form of timely communication. For some types of mechanisms, customers may want or need access to their recent or current usage pattern. In addition, customers' ability to respond to dynamic pricing may be enhanced by control technologies that automate the manner in which the operation of various types of energy-using equipment is modified, as illustrated in the critical price TOU program described in Section 5.2.

There appear to be two main barriers limiting faster penetration of advanced meters. One is the simple cost-benefit question of whether the cost of advanced metering devices and systems can be justified by the benefits that customers and energy suppliers can derive from both cost savings (e.g., replacing manual meter reading) and potential benefits from customers responding to dynamic pricing. This issue has been paramount since the initial studies of time-of-use pricing more than 20 years ago. Metering costs continue to fall, leading some utilities such as Puget Sound Energy and PPL Electric Utilities to justify installing advanced metering systems for nearly all of their customers on cost savings alone.

The other issue is who should and will own the meters, and thus who has the incentive to install advanced meters and recover the cost. The metering function is potentially a competitive service. However, it is also considered a natural extension of the distribution wires system that should be owned by the regulated distribution utility. Until the issue is resolved in each state, no party is willing to make an investment with an uncertain prospect of cost recovery.

Measurement and other implementation issues

As discussed in Section 4, a fundamental issue for the load reduction type of demand response mechanism is measuring the amount of demand response actually provided. Estimation of customers' baseline load and insuring against problems of gaming will remain a potentially contentious issue for both customers and LSEs, as well as the ISO/RTO. In retail competition, the baseline is still an important concept, but the contentiousness largely disappears. Instead, in competitive retail markets the baseline is established through the process of customer choice combined with market-based pricing. Retail customers that want to participate in demand response markets can purchase forward contracts at market-based prices. The baseline quantity, established by the forward contracts, result from customer choice given market prices and price volatility.[37] Furthermore, measuring individual customer load response is not an issue for dynamic pricing. Customers simply pay for what they consume, while risk management (and revenue recovery, for regulated utilities) is accomplished through financial contracts for differences that do not influence behavior at the time that dynamic prices occur. All of this suggests that the baseline load issue will remain a challenge for the load reduction category of demand response mechanisms.

Factors affecting different categories of demand response

Dynamic pricing

Demand response through dynamic pricing can occur in regulated retail markets; however, utilities need some incentive to offer such pricing options. The most successful dynamic pricing program to date in a regulated market is arguably the real-time pricing program offered by Georgia Power Company. This program arose in large part as a response to a limited form of competition in Georgia, as a means for encouraging economic growth at low incremental prices. Though Georgia Power remains a regulated utility, large new loads in the state are allowed to choose suppliers from among Georgia Power or other providers. Georgia Power designed a two-part RTP tariff that charges standard rates for a baseline level of usage, but offers hourly market-based prices on incremental load. Those prices have been lower than standard rates on average, allowing customers to grow at low market prices, but can rise to high levels during periods of system constraints, thus encouraging demand response when it is valuable. Other utilities that have also experimented with two-part RTP include Niagara Mohawk Power Corporation, Duke Power Company, Cinergy, Utilicorp, Xcel (Public Service Company of Colorado), and Kansas City Power and Light.

---

37 This is in contrast to negotiated, mandated, or rule-based determination of baseline quantities that occur in non-competitive retail markets.

The limited extent of dynamic pricing at regulated utilities suggests that some special circumstance needs to be in place to encourage its adoption. Among the special factors that appear to have led to adoption of dynamic pricing are the following:

- Competition for retail load (e.g., Georgia Power),
- Desire to expand demand response to residential customers (e.g., Puget Sound Energy),
- Pressure from, or willingness of regulator to innovate (e.g., Georgia), and
- Internal “champion” (e.g., Bernie Neenan at Niagara Mohawk Power Corporation in the late 1980s and Michael O’Sheasy at Georgia Power in the 1990s).

In general, regulated utilities and their customers need economic incentives to offer and accept retail price structures that involve variable prices and thus price risk. As an example, consider the case in which a utility’s standard tariff is unbundled, with customer charges, and transmission and distribution wires charges set under cost of service regulation, but energy prices based on wholesale market prices—that is, based on future wholesale prices or current expectations about those prices in the form of forward prices. In that case, retail energy prices would be determined using the risk management strategies described in Section 3. That is, a range of price structures would be offered that differ according to the allocation of risk between the customer and the utility. Any fixed prices would include risk premiums to protect the utility against load and price risk. Some customers would voluntarily forego this risk premium, opting to “self insure” by facing lower prices on average, and shifting usage from high-price to low-price periods. Others would opt to face dynamic prices, but would manage price risk through financial contracts for differences.

This type of approach might appear to increase risk to customers. However, the risk imposed by volatile prices cannot be avoided. It must be paid currently through a risk premium on current fixed prices, in the future through balancing accounts, or faced directly in the form of dynamic prices.

Interruptible and voluntary load reduction programs

In the absence of dynamic pricing, and depending on regulatory rules, energy suppliers can have an incentive to pay customers to reduce consumption during periods of high wholesale costs, as illustrated in Section 4.3. These load reductions may take the form of traditional load management programs, or the energy buy-back programs currently offered by some utilities. As a result of the load reductions, the utility or LSE can avoid paying some of the high wholesale prices, or profit by selling freed-up energy into the wholesale market.

It is not clear what incentives ISOs/RTOs would have to operate demand response programs if dynamic pricing on the part of utilities and LSEs became more prevalent through state regulatory actions. However, in the absence of dynamic pricing, an ISO/RTO can have an incentive to operate a demand response program in order to bring price responsiveness to the demand side of the market and improve the market's operation during periods of capacity constraints and/or emergencies. A number of fundamental design issues must be addressed. Key among them are which markets to open to demand response, how much to pay for load reductions, how to measure them, and how to coordinate with the wholesale providers of demand response.

Demand response and ancillary services

As noted earlier, demand response can play a larger role in reserves services markets the longer is the period of response time. Hirst (2000a) discusses many of the technical issues involved in determining the appropriate role of demand response in these markets, and the extent to which current operating restrictions should be relaxed to accommodate customer loads in place of generators. A key question is how many customers will be willing to go to the trouble to bid their load into these markets.

Summary of demand response with non-competitive retail markets

The regulatory and technological barriers discussed in this section have limited the scale of demand response mechanisms in the current transition period. The lack of incentives for, or regulations permitting utilities to offer dynamic retail pricing under traditional cost-of-service state regulation appears to be the greatest barrier to demand response. However, as shown by the success of Georgia Power's two-part RTP program and others, pricing designs are available under a regulated environment that can offer customers dynamic pricing while assuring revenue recovery to the utility and price risk management to customers.

In the absence of price responsive load through dynamic pricing, however, a number of demand response programs have been designed by ISO/RTO organizations. These programs have only been in existence for a couple of years, and appear to have achieved modest success in subscribing customer loads through LSEs. At the same time, a number of issues continue to be debated and program designs modified. Among these are the method for estimating the baseline load, the incentive payment to be paid to customers for load reductions, the problem of lost distribution wires revenue as a result of lower usage levels, and the coordination of programs offered by ISO/RTOs, and utility and LSE contracts with their own customers.

The limited success of some programs has led to suggestions that incentive payments should be increased to encourage greater participation. This paper, however, has pointed to natural market-based mechanisms based on a benchmark of competitive markets that provide a roadmap to greater levels of demand response in retail markets without the need for subsidies or complex rules and regulations.

6.4 Fundamental questions and basic design issues

In Section 2 of this paper we presented goals and objectives of incorporating demand response into market design. We then presented a set of fundamental questions, challenges, and basic design issues that need to be addressed. We now return to those fundamental questions, challenges, and design issues and attempt to give preliminary answers and insights.

Fundamental questions

- *How should demand response be accommodated in wholesale energy, ancillary services, and transmission congestion markets?*

  Demand response can be incorporated in all day-ahead, and hour-ahead electricity markets. In addition, demand response can be incorporated into the reserve services markets with longer response times. The amount of load resources that can effectively participate in each of these markets increases as the allowed response time increases. Only in the real-time balancing market and the regulation and short-term spinning reserves markets is demand response likely to be inadequate because of response time. However, incorporating demand response into the other markets may indirectly impact these short-term markets by increasing the availability of supply-side resources.

- *What is the appropriate role of the RTO/ISO with respect to demand response? Is it to establish rules, facilitate markets, and play the role of lifeguard/policeman? Or should it directly recruit demand response resources through programs such as those offered in New York, New England, PJM, and California?*

  RTOs and ISOs have the general responsibility for providing a non-discriminatory market environment that promotes economic efficiency, maintains power system reliability, lowers the cost of delivered energy, and increases choices offered to wholesale market participants. In addition, the market authority has the duty to monitor market performance and mitigate the exercise of market power. Thus, the RTO/ISO primary role is of market monitor, rather than as market participant. If, however, market performance fails to meet the objectives of efficiency or reliability, direct recruitment of demand resources by the RTO/ISO may be warranted. This is more likely to be the case in the ancillary services markets. Competitive retail markets, with more choices of dynamic pricing products, would reduce the need for RTO/ISO intervention in wholesale markets.

- *What is the appropriate role of both existing utilities and non-utility load-serving entities (LSEs) with respect to demand response?*

  Utilities, or LSEs in restructured markets, are the natural load-serving entity to deliver retail demand response into the wholesale market, because the utilities are the buyers in the wholesale market and they have the retail billing arrangement with customers. To do this, however, utilities must offer retail customers market-based dynamic pricing options. Furthermore, to be effective, these dynamic pricing options cannot be undercut by below-market standard service offers.

Currently, the lack of dynamic pricing choices in retail markets, and regulatory mandated revenue recovery mechanisms have created a role and incentives for non-utility load serving entities (e.g., Curtailment Service Providers) to directly recruit retail customer load response for bidding into wholesale market. Aggregators may improve the performance of the wholesale market by increasing the retail load participation, but in doing so may disrupt the collection of fixed distribution wires costs. The delivery of retail load directly to the wholesale market by aggregators is a contentious issue because some non-participating parties may actually be harmed (have to pay a higher amount of the fixed costs) as a result.

- *Which regulatory jurisdictional questions need to be resolved?*

  Jurisdictional issues are fairly clear. FERC has authority in most wholesale markets, while state regulators have authority over retail rates. Coordination and common interest among regulatory agencies are the needed ingredients to successfully incorporate demand response into wholesale markets. Demand response participation in wholesale markets ultimately depends upon having some retail load on market-based retail pricing. However, state regulatory policies, such as non-market-based standard service offers, cause conflicts in the delivery of retail load into wholesale energy markets.

- *Is there a minimum or optimum amount of demand response? If so, how should it be determined?*

  It is clear that a small amount of demand response would yield significant effects in mitigating price spikes and checking wholesale market power. However, there is no a priori minimum or maximum amount of load response. Instead, what is needed is removal of barriers to market-based pricing and customer choice. Given customer choice and market-based pricing, the optimum amount of load on dynamic pricing structures will be a market-determined result.

Challenges and basic design issues

- *How can demand response be incorporated into wholesale market design when the retail markets that are the source of demand response are in various states of restructuring, including continuing regulation, transition toward retail competition, and near complete deregulation?*

  Two approaches can be simultaneously pursued to incorporate retail demand response into wholesale markets. The first approach is to undertake reform of some of the current regulatory framework that hinders load response. In particular, using market-based pricing principles for the standard service offers (default service, provider of last resort, etc.) would help the success of market-based dynamic pricing offers. Also, modifying the method of collecting distribution costs (which are largely fixed) to a non-volumetric approach (e.g., demand charge or customer-specific access charge) would facilitate more efficient retail energy price signals being conveyed into retail markets.

The second approach to incorporating demand response into wholesale market design while retail markets remain regulated is to encourage LSEs to develop innovative dynamic pricing structures and offer customers greater choice in pricing structures. These pricing structures would allow the status quo contribution to fixed costs to be maintained, while conveying wholesale market information to retail consumers. Two-part RTP pricing, with the baseline load being set at expected (historic) usage and priced at the standard tariff is an example of such a structure.

- *How can demand response be incorporated into market design in a manner that is fair to all parties, where fairness requires that costs be borne by those market participants that cause the costs? Fairness may also require that all market participants have non-discriminatory access to participate in demand response programs and their benefits.*

  Two-part dynamic pricing structures coupled with voluntary customer participation comes the closest to meeting the fairness criterion. All customers would continue to make the same payments to system fixed costs as under their standard tariff. Customers participating in dynamic pricing programs would benefit by exploiting their ability to respond to market prices. All market participants would benefit from improved market performance.

- *How can demand response be incorporated into market design in an economically efficient manner, without cross-subsidies?*

  The most straightforward mechanism in which demand response may be efficiently incorporated in market design is through the combination of market-based pricing and customer choice (from among multiple market-based pricing offers, if not suppliers). Inducing participation through non-market based subsidy payments or mandating a minimum level of load participation can lead to cross-subsidies and inefficient outcomes.

- *What enhancements to technological infrastructures (e.g., metering, communications) are needed to support demand response, and how will the associated costs be recovered?*

  Dynamic pricing structures require hourly interval metering and communication technology to convey wholesale market prices to the customers. The cost of metering and communication technologies has been decreasing. The costs of the needed demand response technologies have typically been recovered by charges to participating customers. Recently, in California and Washington state, installing advanced meters has been financed across all taxpayers or rate payers. While the social benefits of having some retail load on dynamic rate structures might justify some subsidization of installing technology, putting these costs directly on program participants helps insure against cross-subsidization.

7. CONCLUSIONS AND RECOMMENDATIONS

The material in this report may be summarized in a few key points.

1. Demand response is a key element of market design. The near-universal sentiment that encouraging demand response, or price-responsive load, on the part of retail customers is a necessary element of effective wholesale power market design is undeniable. The current lack of demand response leads to a number of problems in otherwise competitive wholesale markets, including wholesale price spikes, reliability problems, possible increased opportunity to exercise market power, and a perceived need for excessive reserve capacity. Two logical categories of demand response in energy markets are *dynamic pricing*, and some form of *load reduction* program, or *demand bidding.*

2. Utilities and LSEs are positioned to provide the natural market mechanism for demand response. In a "normal" market environment of competitive wholesale and retail markets, demand response mechanisms should arise naturally, without a need for non-market incentives, or subsidies. Some customers would willingly choose dynamic retail prices that vary with wholesale costs (especially if they can use financial risk management mechanisms to limit their price risk), and LSEs would have incentives to offer such price structures. These customers would gain access to low wholesale prices during many hours of the year, while during occasional periods of high wholesale costs, they would have a financial incentive to reduce load. Their LSEs would in turn have an incentive to bid price-responsive loads into the wholesale market, thus providing a mechanism for incorporating demand response into the wholesale market.

3. The dynamic pricing category of demand response mechanisms follows most naturally from a competitive market framework. Achieving a greater penetration of dynamic pricing will require encouragement from state regulators for utilities to expand their standard tariffs to offer customers a choice from a range of price structures, including some that provide *dynamic pricing.* Available pricing strategies (e.g., two-part RTP for large commercial and industrial customers, and critical price TOU for residential and small commercial customers) can give customers dynamic pricing incentives, while still meeting regulatory concerns about revenue recovery and price risk. These retail-pricing programs provide a natural link between the wholesale and retail markets, and by acting in place of a competitive retail market allow an efficient pricing mechanism for markets in transition. A potential barrier to making dynamic prices available to all customer types is the lack of advanced metering devices needed to record usage and communicate prices.

Dynamic pricing carries certain advantages relative to load reduction, or demand-bidding programs, including the following:

- Customers pay for what they consume, at prices tied directly to wholesale market costs; there is no issue of measuring load reductions.
- The demand response is provided through the customers' regular contract with their energy provider.
- Customers can focus on their primary activity of producing a product or providing a service, rather than deciding whether and when to sell energy into wholesale power markets.

While individual customer response to dynamic pricing may be considered "passive," the total amount of load response by dynamic pricing customers can be forecast by their utility or LSE, who will have an incentive to do so in order to schedule load accurately in the wholesale market.

4. The other category of demand response mechanisms—load reduction, or demand bidding programs—face a number of design challenges. The current environment of largely regulated retail prices and little dynamic pricing arguably creates an apparent need for ISO/RTO market intervention to encourage some form of demand response. The approach being followed in PJM, New York, and New England is *demand bidding programs* offered through the ISO/RTO, in which customers or load aggregators bid load reductions into daily market auctions along with generators. Several key questions arise in the design of such programs. Some of those questions, with responses, include the following:

- *What payments or incentives should be paid for load reductions?* A review of the incentives provided by competitive markets suggests that suppliers should be willing to pay no more than the wholesale cost of power for load reductions from a baseline load *for which customers have paid the retail energy price.*[38] Any higher payment implies the need for some form of subsidy to cover the difference between the revenue paid out in the incentive payment and the cost saved by the load reduction. Naturally, the larger the incentive payment, the larger will be customers' demand response; however, the market price provides the indicator of the appropriate value and amount of load reduction.

---

38 This is algebraically equivalent to a condition that the supplier should be willing to pay no more than the difference between the wholesale cost and the retail price.

- *How are load reductions measured (i.e., how the baseline load is calculated)?* No single best approach exists to accurately calculate each customer's baseline load during a demand response period and avoid issues of customers "gaming" their baseline load. A competitive market approach suggests the use of pre-determined baseline loads along the lines of forward contracts for fixed quantities, which is analogous to the approach used in two-part real-time pricing.

- How do the ISO program terms relate to customers' contracts with suppliers?

- How many retail customers want to act as pseudo-generators, deciding each day how much load reduction to bid into the wholesale market?

5. The optimal amount of demand response should be determined by the market rather than specified a priori. This is a corollary to the previous conclusion. If market prices signal the value of power, then customers' response to those prices reflects the appropriate amount of demand response—low levels at low prices, and high levels at high prices. Possible qualifications to this conclusion are potential barriers to some customers of having the opportunity of choosing dynamic pricing, such as that provided by the lack of advanced metering.

6. Market designers should not ignore the benefits from the mirror image of load reductions at high wholesale prices—load increases during frequent periods of low wholesale prices. This topic, which receives little attention in most discussions of demand response, may be the source of substantial benefits from market efficiency. However, the benefits from demand response to low prices will likely only be achieved through dynamic pricing, as demand bidding and load management programs are typically designed to operate only during infrequent periods of high wholesale prices.

APPENDIX

CUSTOMERS' BILLS UNDER TWO-PART RTP

The customer's bill under two-part RTP may be expressed in two alternative but equivalent forms. Each can be useful in illustrating certain features of two-part RTP. The first form focuses on the revenue neutrality of two-part RTP relative to the standard retail tariff, or a contracted fixed price, for the amount and pattern of energy use represented by the customer's baseline load (CBL):

$$\text{RTP Bill} = \textit{Base bill} + \textit{Sum}\,[P^R_h * (E_h - CBL_h)].$$

The *Base bill* is calculated as the customer's CBL billed at the standard tariff, P^R_h is the RTP price in hour *h*, E_h and CBL_h represent the customer's actual load and CBL in that hour, and *Sum* represents the summation over all hours in the month. This form of the bill shows that two-part RTP is revenue neutral for each customer relative to his CBL. If he maintains his usage at the CBL level in each hour (i.e., $E_h = CBL_h$), then his bill remains equal to what it would be under the standard tariff. In any hour in which his usage under RTP exceeds his CBL (i.e., $E_h > CBL_h$), then he pays a *charge* equal to the incremental usage times the RTP price. Alternatively, whenever he reduces usage below the CBL level under RTP, then he receives a *credit*, valued at the RTP price.

The actual form of the *Base bill* may be somewhat complicated depending on the form of the standard tariff (e.g., it may include demand and energy charges). However, conceptually the *Base bill* may be converted to the sum over the hours in the month of an hourly effective energy charge under the base tariff, P^B_h, multiplied by the CBL, or *Sum* (P^B_h * CBLh). Substituting this term into the above equation, and rearranging terms, produces the following equivalent form for the RTP bill:

$$\text{RTP Bill} = \textit{Sum}\,(P^R_h * E_h) + \textit{Sum}\,[(P^B_h - P^R_h)* CBL_h].$$

This form of the bill illustrates the customer's incentives to respond to RTP prices, and the risk management feature provided by the CBL. The first term shows that under two-part RTP the customer's *entire load* is exposed to RTP prices. That is, any change in consumption in response to RTP prices, regardless of whether usage is above or below the CBL, has a direct impact on the bill. The second term can be characterized as an access charge that illustrates the risk management feature provided by the CBL; when RTP prices are high (e.g., $P^B_h < P^R_h$), the access charge during those hours is negative, thus partially offsetting the effect of the high RTP prices on the customer's bill. Note that the customer has no control over the factors in the second term; that term represents purely a financial adjustment, a *contract for differences*. The customer's only opportunity to affect his bill is to modify usage in response to the hourly RTP prices.

The above discussion has used the example of two-part RTP offered by a regulated utility that has the authority to recover revenue from customers through a standard tariff. However, the same bill equations can be used to illustrate how two-part RTP would operate in a competitive environment. Focusing on the second form of the bill, a customer could agree to purchase all of its energy from a supplier at RTP prices that are based on wholesale market costs. Then he could enter a separate financial contract for differences agreement to protect a portion of his load, represented by the CBL, at a forward price set by the market, P^B_h. The key differences would be that the customer would choose the amount of the CBL, rather than having it set by the utility at historical levels, and the price for the CBL would be set by the market rather than the regulator.

REFERENCES

American Electric Power 1992, *Report on the Variable Energy Pricing and TranstexT Advanced Energy Management Pilot.*

C. Aubin, D. Fougere, E. Husson, and M. Ivaldi 1995, "Real-Time Pricing of Electricity for Residential Customers: Econometric Analysis of an Experiment," *Journal of Applied Econometrics,* 10, S171–S191.

S. Braithwait 2000, "Residential TOU Price Response in the Presence of Interactive Communications Equipment," in *Pricing in Competitive Electricity Markets*, edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, Boston, MA.

S. Braithwait and A. Faruqui 2001, "The Choice Not to Buy: Energy Savings and Policy Alternatives for Demand Response," *Public Utilities Fortnightly*, March 15.

S. Braithwait and M. O'Sheasy 2002, "RTP Customer Demand Response—Empirical Evidence on How Much Can You Expect," in *Electricity Pricing in the Transition*, edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, Boston, MA.

EPRI, Palo Alto, CA: 2001. 1005945. *Customer Response to Electricity Prices: Information to Support Wholesale Price Forcasting and Market Analysis.*

D. Caves, K. Eakin, and A. Faruqui, "Mitigating Price Spikes in Wholesale Markets through Market-Based Pricing in Retail Markets," *The Electricity Journal*, April 2000.

K. Eakin and A. Faruqui 2000, "Pricing Retail Electricity: Making Money Selling a Commodity," Chapter 2 in *Pricing in Competitive Electricity Markets*, edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, Boston, MA.

Federal Energy Regulatory Commission 2002, *Working Paper on Standardized Transmission Service and Wholesale Electric Market Design*, March.

Federal Energy Regulatory Commission 2002b, *Standard Market Design Notice of Proposed Rulemaking,* July 31.

J. D. Glyer 2000, "The Price of Electricity: What the Contract Says Doesn't Necessarily Reflect the Customer's Incentives and Perceptions," EPRI International Energy Pricing Conference, July.

C. Goldman, G. Heffner, and G. Barbose 2002, *Customer Load Participation in Wholesale Markets: Summer 2001 Results, Lessons Learned and "Best Practices,"* Lawrence Berkeley Laboratory, February.

E. Hirst 2002a, *Price-Responsive Demand as Reliability Resources, Edison Electric Institute*, April.

E. Hirst 2002b, *Barriers to Price-Responsive Demand in Wholesale Electricity Markets*, Edison Electric Institute, June.

E. Hirst and B. Kirby 2001, *Retail-Load Participation in Competitive Wholesale Electricity Markets,* Edison Electric Institute, January.

W. Hogan 2002, "Electricity Market Restructuring: Reforms of Reforms," *Journal of Regulatory Economics*, 21(1), 103–132.

A. Kahn 2002, "The Adequacy of Prospective Returns on Generation Investments under Price Control Mechanisms," *The Electricity Journal,* March.

L. D. Kirsch and R. C. Hemphill 2000, *Redesigning Distribution Tariffs for Restructured Electric Power Markets,* Edison Electric Institute, April.

Neenan Associates, LLC, "-NYISO Price-Responsive Load Program Evaluation Final Report, January 8.

C. Rochlin 2002, "Is Electricity a Right?" *The Electricity Journal*, March.

L. Ruff 2002, *Economic Principles of Demand Response in Electricity*, Edison Electric Institute, September.

EDISON ELECTRIC
INSTITUTE

701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000
www.eei.org