Demand-Side Resources and Regional Power Markets: A Roadmap for FERC

Richard Cowart

Draft 2A – 1/04/02*[*]

Introduction

Customer-controlled resources can play a crucial role in creating efficient regional power markets, lowering price volatility and generator market power, disciplining power costs, and improving reliability. Participants in the RTO Futures process have asked for a "roadmap" of actions that FERC could take to advance development of those resources. This white paper sets out that roadmap and sets the stage for discussion by RTO Futures members and their advisors.

Section I: Overview and Summary:

Why should FERC concern itself with demand management and demand-side resources? As an agency focused on electric transmission wholesale power markets, demand-side resources have historically not been an active part of FERC's docket or policy initiatives. This paper is motivated by three fundamental changes to that historic situation: (1) the increasing significance of regional wholesale power markets and regional open-access transmission systems; (2) the breakup of vertically integrated franchises, which has weakened the role of integrated resource management at the retail level; and (3) the development of multi-function RTOs, under FERC's active direction and supervision. Together, these changes call for FERC's active engagement to ensure that the benefits of customer-controlled resources can be delivered to power systems and markets.

Capturing those benefits will not be simple or automatic. Utilities and state utility regulators learned during the 1980's that they needed to pursue demand-side resources in a variety of policies and venues – e.g., through rate design, in siting and certificate of need reviews, in utility IRP and DSM plans, interruptible contracts, and so on. Similarly, FERC will need to explore a number of venues and policies to bring the multiple benefits of demand-side resources to today's regional power systems and markets.

Overall recommendation:

FERC should initiate a thorough review of its authority, responsibility, and opportunities to promote cost-effective demand-side resources in the markets and power

[*] Prepared for the Regional Power Working Group by Richard Cowart, Director, Regulatory Assistance Project. Several other members of the Working Group, notably Sue Tierney and Terry Black, provided helpful input, but the errors are entirely the author's. This discussion draft has not been reviewed or approved by the Working Group, or any of its members.

systems subject to its jurisdiction, and should adopt market structures, transmission tariffs, and RTO policies designed to bring those resources forward.

The important FERC actions are as follows:

1. In order to reveal an economic demand curve as well as a supply curve, FERC should require the operators of wholesale power markets to incorporate demand-side bidding in their market designs.

2. Power market design should accommodate price-responsive load and demand resales, through market rules that give willing customers and their suppliers reasonable opportunity to adjust consumption in response to market conditions. FERC should require adoption of multi-settlement systems or similar market rules to meet this need. FERC should also require RTOs to adopt and invest in standard real-time information systems and communications protocols that may be used by innovative energy service providers and load-serving entities to manage demand-side resources in response to market conditions.

3. Where wholesale power costs are assigned to load-serving entities based upon customer class load profiles, FERC should require market rules that make alternative load profiles available for major load subclasses based upon energy use patterns and load control options.

4. FERC should require that market rules allow demand-side resources to supply ancillary services on an equal basis with supply-side resources. The criteria for supplying ancillary services should be written in technology-neutral terms, and should not require costly telemetry and metering for individual demand-side resources whose performance could be verified on a statistical, aggregated basis.

5. RTOs, reliability managers, and transmission owners often seek cost recovery in FERC-approved tariffs for investments intended to enhance system reliability. Before granting recovery that would broadly "socialize" those costs, FERC should require applicants to show that the benefits are broadly dispersed, and that they have selected the lowest-cost resource, including demand-side resources, reasonably-available to meet the need in question.

6. Transmission prices that hide from customers the costs of congestion, and the value of congestion relief, diminish the reliability contribution that could be made cost-effectively by load management, efficiency, and generation in load pockets. FERC should promote transmission rate designs that reveal the cost of congestion across different times and locations.

7. FERC should encourage RTOs to examine region-wide, reliability-enhancing investments in demand-side resources that would improve reliability and lower power costs. When supported by cost-effectiveness analysis, RTOs should be permitted to recover those investments on the same basis as regional transmission investments,

ancillary service costs, or other RTO expenses.

8. The reliability and economic benefits of demand-side resources will be best developed where wholesale market rules and retail pricing and demand-bidding rules are consistent. FERC should encourage efforts within each RTO to bring state and federal regulators and interested stakeholders together to identify unintended conflicts and to develop a coherent set of market rules, tariffs, and policies in support of cost-effective distributed and demand-side resources.

Section II. Power Market Changes and Consequences

The U.S. electric system is in the midst of a transformation as dramatic as any it has experienced since the emergence of the franchise system early in the last century. The nation is now dealing with the consequences of this transformation, not all of them anticipated by the advocates of reform. One crucial aspect of the new industry structure, and one that was little anticipated, is the growing significance of regional power institutions -- Independent System Operators, Regional Transmission Organizations, Transcos, and Power Exchanges. (There are many important differences among them, but for simplicity here we will refer to them all as RTOs.)

Their growing significance is the result of two complementary policy thrusts. First, FERC, in its open access orders and RTO initiatives, has created new responsibilities for transmission owners in support of broader regional power markets. And, second, state restructuring statutes have purposefully limited the role of traditional franchise utilities, while FERC has encouraged the development of wholesale markets and market-based power transactions, shifting the locus of power supply decision-making to the regional wholesale level.

The consequence of these two trends has been the emergence of new regional power entities with vastly increased influence over power markets, but without the broad regulatory oversight and operational traditions, and without the "least-cost" mandates that guided operation of their predecessor institutions.

The increasing importance of regional power entities has been driven by policy and structural changes at both the federal and state levels. At the federal level, the principal agent has been FERC. Since the passage of EPACT and the move to open access in natural gas pipelines, FERC has taken dramatic steps to open up the electric transmission grid to competing power providers, and to establish RTOs in relevant power markets across the nation. EPACT and FERC's initiatives both recognize and drive the evolution of new regional power entities, including RTOs, ISOs, Transcos, and power exchanges.

At the same time, many states are restructuring their electric franchise systems, among other things transferring functions that formerly occurred ***within*** state-regulated franchises to functions that will be regulated, if at all, at the ***federal, wholesale*** level.

A. Increasing Significance of ISOs and Regional Power Markets:

As former FERC Chair James Hoecker stated last year, “RTOs are no longer a policy possibility -- they are a practical necessity.” FERC’s actions in recent months have only underscored the Commission’s continuing commitment to the RTO concept. The increasing significance of RTOs and new regional power markets can be seen in numerous indicators, including:

> > decreasing reserve margins, increasing reliability challenges, and increasing calls for new reliability rules and institutions such as NAERO to make reliability standards mandatory across very large regions; (according to NERC, EPRI and others, large regions of the country are closer to the edge of unreliable supply than they have been at any time in the past 35 years.)
> > the increasing percentage of power transactions occurring at the wholesale level;
> > the rapidly increasing dollar flows accompanying those transactions;
> > the large number of retail customers in states where traditional franchise obligations have been altered by divestiture and/or the move to potential retail competition;
> > the increasing role of transmission transactions (up 400% in the past 4 years); and the impact of transmission tariffs and access rules on renewable and distributed resources;
> > the provision of ancillary services through regional markets; and
> > the emergence of regional power exchanges and trading hubs, with market rules that have dramatic impacts on the level and volatility of power costs across broad regions.

B. Weaker Franchises and the Demise of IRP:

The dramatic transfer of authority to RTOs has been magnified by actions of many states in restructuring legislation and orders that have greatly reduced the authority and responsibility of incumbent franchise utilities. During the 1980s and into the 1990s, utilities in almost every state engaged in the process of Integrated Resource Planning (IRP), a purposeful process of analysis and management to provide for customers’ energy needs through a least-cost combination of supply-side and demand-side resources. Many utilities demonstrated during that period that customer-based energy efficiency and load management could provide large reservoirs of low-cost resources to the electric system.

In the pursuit of other restructuring goals, much progress on the demand side has now been lost. In many cases, power supply and portfolio maintenance are no longer the responsibility of the historic utility, at least not in the same way as when it had the traditional, long-term obligation to serve. Thus, portfolio management tasks -- such as finding a balance between long-term and short-term power sources, balancing load and generation geographically, attaining fuel diversity, balancing supply and demand-side resources, and investing in energy efficiency -- are no longer performed by those historic franchise utilities. In almost all states, incumbent utilities inherited the responsibility and opportunity to serve default, non-choosing customers. By law (e.g., California) or by choice due to market uncertainties, most default service providers do not engage in least-cost supply planning. Divestiture of generation deprives the franchise service provider of the opportunity to manage fuel, price, environmental, and demand/supply risk for the benefit of customers within an owned portfolio.

Franchise utilities also face new conditions with respect to responsibility for transmission adequacy, and their ability to deliver consumer, environmental, and other public benefits to the nation. Default service regulations, wires company rate designs, and market risks due to potential loss of load have all encouraged the remaining wires companies to abandon the practice of integrated resource planning and are major barriers to the continued acquisition of energy efficiency and other low-polluting resources. Many legislators and regulators simply assumed that these public-purpose functions would be supplanted by market forces. Utility spending on energy efficiency programs has been cut in half nationwide, while faster-than-expected load growth, price spikes, near-term capacity shortages, and price rises have threatened both reliability and the political support necessary to sustain the transition to competitive electricity markets.[1]

C. The Crucial Missing Link: Harmonizing Wholesale and Retail Market Rules to Support Cost-effective Load Management and Efficiency Resources

Following the price explosions and reliability meltdowns in the California electricity markets last year, it has become increasingly obvious that wholesale market structures in California and the West need serious renovation. Regional power markets in other regions are also in need of development and reform. Many observers also understand that retail market rules may be contributing to the problems that are now surfacing. What seems to be less well understood is the critical nature of the ***relationship between the retail and wholesale markets.***

To give one example: if the theory of the wholesale market is that a single-price auction will efficiently reveal the market price of power at each hour of the day, then it is essential that that signal be meaningfully communicated to retail customers (or at least to their providers), so that price can be formed as a function of both supply *and* demand, not just supply. Thus, load-serving entities (LSEs) that skillfully manage load to lower power costs in the market should see the benefits of their management in the form of lower power costs. (And the markets should be structured so that customers see the benefit of those lower costs as well.)

To do this, we have to create market rules that create viable business opportunities for price-sensitive, load-managing LSEs. Unfortunately, current wholesale market rules stand as a wall between wholesale power costs and retail market opportunities. At the simplest level, the load profiles used almost everywhere at the wholesale level to assign time-of-use consumption responsibility to LSEs do not now reward those LSEs who achieve better than average load characteristics.[2] And the market rules in most regions make it very difficult to capture, bid-back, and reward the value of load-side curtailments at high-priced peak periods. The price and demand characteristics of wholesale electricity markets will be distorted unless LSEs and retail customers see the cost of power and the value of load management, and can communicate through demand-side bids the value of consumption and the price required to manage load.

[1] A thorough review of reliability and market challenges, and the potential role of demand-side resources in meeting them, can be found in R.Cowart, "Efficient Reliability: The Critical Role of Demand-Side Resources in Power Systems and Markets" (NARUC, June 2001).

[2] Load profiles that treat all customers in a large class the same assume that they all have the same patterns of consumption. For example, if an LSE who cycles off residential air conditioners during high summer peaks is charged the same as an LSE who has an uncontrolled high peak-hour demand, the average load profile will nullify the wholesale market's pricing signals. This problem is discussed in more detail in Section IV, below.

Load management is not the only demand-side resource impaired by these new conditions. Energy efficiency investments are also undermined by current utility structures and market rules. Those who manage and regulate regional power markets need to be aware of the significant benefits that improved efficiency brings to regional power systems and markets. Investments in energy efficiency can improve system reliability, lower market clearing prices, and lower the overall cost of power for all customers in a power market. Cost-effective energy efficiency investments, deployed broadly in the economy, are modular, dispersed, low-cost, and "dispatchable" in the same hours as the underlying load they are associated with. They lower throughput, fuel consumption, and stress on the transmission and distribution system at all hours, including peak hours. Energy efficiency investments are an essential complement to load management initiatives – both should be seen by decisionmakers as critical to the success of electricity markets in the coming decade.

Energy efficiency, like load management, lowers peak demand and contributes to system reliability. But the high value of demand reduction is not limited to a few peak hours. Studies performed on wwholesale power markets in California, New England, and the PJM region reveal that efficiency's load reduction during all hours of the year reduces the market price of power by many times the cost of the efficiency programs. In PJM in the year 2000, a five percent reduction in load would have produced lower wholesale power costs of an average value of 25 cents per kwh during summer afternoons and 3.5 to 6 cents per kwh in the off-peak and winter periods.[3]

Massachusetts provides another example of the new value of energy efficiency in regional power markets. The Massachusetts Department of Energy Resources concluded that the measures installed in their post-restructuring efficiency program lowered participants' electricity costs by $20 million in 1999. But they also concluded that, by lowering peak demand at high cost periods, the programs provided reliability benefits and power cost savings to all customers, participants and non-participants alike. The benefit to other Massachusetts customers in just 13 hours on one high-cost day exceeded $6 million. Benefits to customers outside of Massachusetts were not calculated.

The regional reliability and wholesale power cost advantages of energy efficiency are squarely within the purposes of FERC's regulation of power markets, reliability rules, and RTOs as set out in the Federal Power Act and in FERC orders. Reliability and market rules should be crafted with a view to advance efficiency investments throughout the region through a range of mechanisms available to FERC and the wholesale and transmission entities that FERC regulates.

The essential point is that the wholesale and retail markets are two parts of the same value chain, and they must be properly linked. ***This raises a particular challenge for decision-makers, since FERC does not have jurisdiction over retail market rules and states do not have jurisdiction over wholesale market structures.*** New linkages have to be forged. If regional power markets and transmission rates are to be sensibly structured, FERC cannot ignore this vacuum, but must

[3] This is calculated as the reduction in wholesale prices due to the demand reduction, multiplied by the energy bought in the wholesale market, divided by the amount of the load reduction achieved. Mid-Atlantic States Cost Curve Analysis (Marcus 2000).

initiate policies that would support and encourage efficient decision-making by state decision-makers, transmission providers, and market participants.

Section III. FERC in Action -- Policy Opportunities in Current FERC Initiatives

It is becoming clear that a demand response has enormous value to consumers. The problem is that the structure of our markets hides that value from consumers, retail sellers, and innovative entrepreneurs who could develop new demand-side technologies and services. With the value of demand response so well hidden from those who could profit from it, it is no wonder that we see little demand response in these markets today. Regulators need to structure the market and market rules so customers, retail sellers, distribution utilities, and current and potential vendors of demand response have an opportunity to realize the value of the services they can offer.[4]

Fortunately, FERC has begun to focus on this challenge, observing in July, 2000 that "... a successful transition to competitive electricity markets will necessarily involve an increased participation of the demand side of the market in making resource decisions" and that demand-side participation "can serve to discipline prices by bringing supply and demand into balance."[5] Early in 2001, in response to California's power market problems, FERC staff recommended specific actions to facilitate effective demand response in wholesale markets.

In its March 2001 recommendations on the California wholesale power market, the Commission Staff noted that "While demand response is normally thought to be an issue of design, for truly robust demand responsiveness investments in infrastructure are needed." (p. 1) In addition, Staff stated that "Because of the limited demand response in today's California markets, small quantities of supply withheld from the market can lead to very high prices and create strong incentives to withhold supply unless the price received is well above a competitive market response." (p. 18) After recommending that California LSEs be required to identify curtailable load and bid the capacity of that load into the market, along with the price at which the load would be willing to be curtailed, Staff stated that "The ISO should work to provide a mechanism for these demand bids to be implemented as part of the ancillary services market." (p. 23) In sum, the Staff concluded:

> [D]evelopment of the demand side of the power market is critical, and cannot take place without investment in an infrastructure that enables customers to see and respond to price signals. Part of the needed development will be in information and control technology that enables load to respond, but institutional development in the form of rules and protocols for demand participation in the market are needed as well. (p. 28)[6]

---

[4]Of course, not all consumers will want to have to see real-time prices. But many efficiency and load-management programs are essentially automatic from the customer's point of view, and it is appropriate to give consumers the option to see real-time prices, and to enroll with LSEs who will manage load in response to them.
For a summary of this analysis, see Regulatory Assistance Project, "Using A Demand Response to Stabilize Electric Markets" (February, 2001).
[5]FERC Order in NSTAR v. NEPOOL (July 26, 2000).
[6] Staff Recommendation on Prospective Market Monitoring and Mitigation for the California Wholesale Electric Power Market, San Diego Gas & Electric Co. v. Sellers of Energy and Ancillary Service Into Markets Operated by the CA ISO & Power Exchange, Docket Nos. EL00-95-012, *et al*. (March 2001)

To help address problems in Western energy markets this Spring, the FERC ordered that a number of interim actions be taken by the California ISO and other regulated utilities, including actions to facilitate greater demand responsiveness. Among other things, the Commission waived several notice requirements related to back-up and self generation, authorized wholesale (and retail, where permitted under state law) customers to resell their load reductions at market-based rates into the wholesale market, and enhanced DSM cost recovery treatment. In taking this action, the Commission stated:

> It is widely accepted that dropping even a few megawatts off the system at peak periods is more efficient and economical than the incremental cost of generating them. Demand reduction offers a short-term and cost-effective means to provide additional resources during times of scarcity. Therefore, the Commission will allow…retail customers, as permitted by state laws and regulations, and wholesale customers to reduce consumption for the purpose of reselling their load reduction at wholesale. By providing additional load resources when generating resources are scarce, these "negawatts" should help maintain the reliability of the grid. To stimulate the development of this program, the Commission is granting a blanket authorization to allow these sales at market-based rates….
>
> These transactions are considered wholesale when they involve the sale for resale of energy that would ordinarily be consumed by the reseller. These transactions can occur in several ways. An aggregator can line up retail load to acquire enough negawatts to resell in a manner similar to what aggregators do when they sell power to retail load under retail choice programs. In addition, wholesale and retail load with contract demand service could resell their contract demands if the value of power is greater than the value of consumption.[7]

In a subsequent order the Commission implemented several FERC Staff recommendations. It ordered all public utilities purchasing electricity in the CA ISO's real-time market to submit demand-side bids, indicating the prices at which loads would be curtailed and identifying the loads to be curtailed, and it required the ISO to curtail service to the entity in accordance with its bids. As the Order stated,

> These requirements will develop demand-side price responsiveness that will help mitigate market power and lessen the severity of price spikes. When demand responds to price, suppliers have additional incentives to keep bids close to their marginal production costs, because high bids are more likely to reduce the bidder's energy sales. Thus, demand-side bidding applies downward market pressure on prices. Demand-side price-responsive bids will also help to allocate scarce supplies efficiently. Without the development of price-responsive bids, the allocation of short supplies – through rolling blackouts – is arbitrary and inefficient. In order for the market to function effectively, there must be a mechanism to allocate short supplies to those who value energy the most, while encouraging those with lower-cost alternatives to take advantage of them.

[7] Order Removing Obstacles to Increased Electric Generation and Natural Gas Supply in the Western United States, 94 FERC ¶61,272, Docket No. EL01-47-000, March 14, 2001.

> Customers need to be able to respond to price signals so that those facing more elastic demands can relinquish power to those placing greater value on obtaining power at that time. For example, a load serving entity serving a retail customer with back-up power needs to have appropriate price signals to determine whether the back-up source should be used. An industrial plant also could agree to close during certain hours, or blocks of hours, during the day, allowing its load serving entity to reduce real-time purchases from the ISO.
>
> Demand response can also be developed by establishing a western-wide program under which energy users (such as industrial plants) outside of California could be paid for curtailing power to be used in California. This could be accomplished by having the customer voluntarily submit a bid for demand reduction (or interruption), and, if that bid is accepted, the customer would have its power transferred to a deficit control area, and be paid for their load curtailment.[8]

The Commission also observed:

> State authorities can promote demand-side price responsiveness in several ways, such as allowing retail rates to vary to reflect wholesale prices, facilitating the necessary metering, and adopting conservation programs. While the design of retail rates is a matter of state jurisdiction, the requirements adopted here do not intrude upon state retail rate design. Instead, they bear upon the development of prices in the ISO's markets and the rules governing how sellers and buyers act in those markets, over which the Commission has jurisdiction.
>
> The Commission has concluded that it is necessary to require public utility load serving entities to submit demand bids and that demand side bidding should begin June 1, 2001. Although retail demand response may not be fully developed by that time, there are some efforts in effect now and this requirement will support those efforts. The Commission fully expects that price responsiveness of load serving entities will increase over time as retail programs develop and additional metering is installed to allow retail customers to respond to prices. The wholesale requirement for demand side bidding will, therefore, be in place to support those efforts. Moreover, as discussed above, requiring demand side bidding will provide downward pressure on wholesale prices since sellers will recognize the ISO will not pay any price to obtain power.[9]

More recently, the Commission has approved both emergency and economic customer demand response pilot programs for the New England, New York, and PJM ISOs, noting in the PJM order, for example, that "Price-responsive demand is a key part of a well-functioning market that would mitigate price volatility and enhance reliability in the face of supply shortages."[10] The PJM Order also observes:

> In a well-functioning, competitive electricity market, high prices are a signal for buyers

---

[8] Order Establishing Prospective Mitigation and Monitoring Plan for California Wholesale Electric Markets, 95 FERC ¶61,115, Docket No. EL00-95-012, April 26, 2001.
[9] *Id.*
[10] Order Accepting Tariff Sheets as Modified, 95 FERC ¶61306, Docket No. ER01-1671-000, May 30, 2001.

> to conserve and for sellers to expand output. The market would thus allocate scarce energy and capacity to those who valued it most and assure that the load was served at least cost. But, the market structure that has developed in PJM as in the rest of the nation does not communicate wholesale prices to retail customers in real time. Because efficient real time prices are not conveyed to retail customers, they have no incentive to reduce consumption voluntarily to alleviate power shortages.
>
> PJM's proposed Load Response Program…would, in part, address this market flaw. Absent implementation…end users who do not already participate in an LSE's load reduction program may have little or no incentive voluntarily to curtail use at times of scarce supply, since they would receive little or no economic benefit from such curtailment.[11]

In response to parties suggesting that the FERC does not have jurisdiction to approve load response programs, the Commission noted in the PJM Order that load response transactions are "wholesale" when they involve sale for resale of energy that would ordinarily be consumed by a retail customer, citing its Western Markets orders and observing that no state commission in the proceeding asserted that the Commission lacks jurisdiction. Further, the Order concludes:

> In the instant case, the current lack of meaningful demand side response is a flaw in the markets operated by PJM which, if not corrected, could lead to dysfunction in those markets, and the Load Response Program is part of PJM's attempt to correct that dysfunction. PJM's markets are within our jurisdiction, and the Load Response Program is thus within our jurisdiction as well.[12]

In the months ahead, one of the critical tasks related to making wholesale power markets workably competitive is assuring that effective customer demand response is built into the design of regional power markets. In pending RTO dockets, as well as pilot ISO load response program dockets, the FERC will have many opportunities to implement policies that require development of effective demand response mechanisms and protocols. Because of FERC jurisdiction under the FPA, it is the only regulatory body with authority over the design and operation of regional power markets and the institutions that run them. Thus, FERC policies on demand-side resources and their role in regional power markets will determine whether customer demand response becomes an effective part of the electricity marketplace.

Section IV. FERC Demand-Side Roadmap – Eight Policy Recommendations

The focus of most decision-makers on supply-side solutions to meet load growth and reliability needs is perhaps a natural product of the manner in which franchises and electricity markets have evolved. Nevertheless, the demand side of these markets represents an enormous reservoir of untapped capacity available both as a moderating influence in the volatile wholesale electricity

[11] *Id.*
[12] *Id.*

markets and as a resource in its own right in the markets for reliability services. Tapping these resources requires decision-makers to focus on two simply-stated goals. Every effort should be made:

- **to expose the value of demand response and energy efficiency in the wholesale and retail markets to as many participants as possible, and**
- **to identify and remove unnecessary barriers to demand-side resources, and unnecessary preferences and subsidies for supply-side solutions to meet energy supply and reliability objectives.**

Eight policy initiatives for accomplishing these purposes are described below.

1. <u>Demand-Side Bidding</u>: In order to reveal an economic demand curve as well as a supply curve, FERC should require the operators of wholesale power markets to incorporate demand-side bidding in their market designs.

Under traditional franchise regulation, the financial relationship between electric demand and the cost of supply was indirect at best. In particular, the cost of maintaining reliability at peak was rarely reflected in peak period prices. The price signals delivered to both wholesale and retail customers were averaged over time and location, and bore only a general relationship to the cost of production. Supply was managed, not to match marginal cost and marginal "demand" from customers, but to meet the revealed load curve of customers who received only broadly averaged price signals. The "demand curve" in such circumstances was an engineering concept as well as a legal requirement (i.e., the obligation to serve without regard to the cost of supply) more than the revealed willingness-to-pay of the utility's customers. This is the vertical "assumed" demand curve shown in Figure 1.

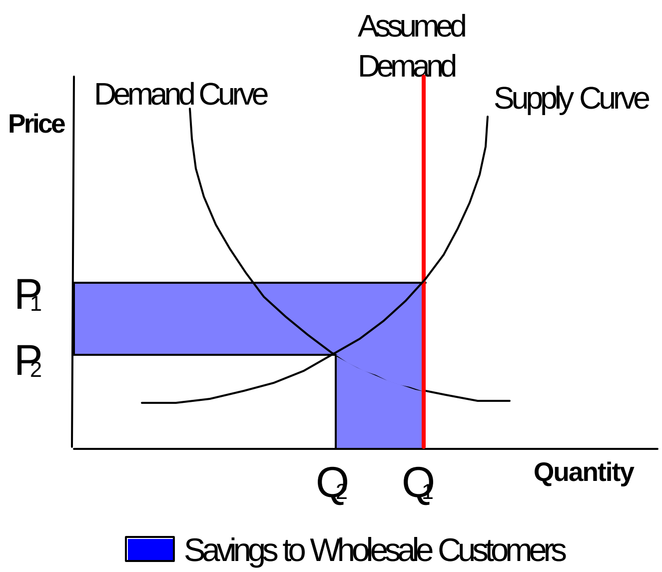

Figure 1: Assumed vs. Actual Demand Curves for Electricity. Revealing the actual demand curve would lower price and peak loads in the market.

While the supply side of the supply/demand equation has changed in most regions of the country, the demand side of the equation has not, to any significant degree. There is still a vertical demand curve. Whatever the merits of this muted demand response in the franchise system, it has serious detrimental effects in electricity markets that are being established to efficiently balance demand and supply. As recent price spikes, high prices, and reliability challenges of those markets reveal, efficient energy markets simply cannot be built on such a foundation. There is ample evidence that for many customers, demand for electricity is moderately elastic, and that at the high prices experienced in tight market situations, customers with choices will respond by reducing demand, or shifting it to hours when prices are lower. Revealing the customers' real demand curve is now a critical challenge for the nation's electric policymakers. As shown in the second curve in Figure 1, markets will clear at lower quantities and lower prices when this curve is exposed.[13]

The good news in this connection is that a relatively modest demand response during just a few hours of the year could make a very substantial contribution to lowering peak demand, and thus enhancing reliability -- and lowering overall power costs at the same time. In New England, for example, 9% of the capacity of the generating system is called on just 1% of the hours of the year, and the high price of power in the most expensive 1% of all hours accounts for 16% of the total annual dollars spent in the spot market.

Another important feature of electricity markets is that a very small fraction of the customer base accounts for a large fraction of total demand. Nationwide, the 538,000 industrial customers represent just 0.4% of total customers, but 30% of total demand,[14] equivalent to about 200 GW of demand at the time of summer peak. This is true for individual utilities as well; 1% of the customer base might well account for 50% of a utility's load.[15]

For those constructing wholesale electricity markets, the lessons suggested by these data are relatively straightforward: First, as Hirst and Kirby conclude:

> These results show that large benefits can likely be achieved by offering dynamic pricing and related programs to a tiny fraction of the nation's electricity consumers. In other words, your grandmother need not worry about responding to real-time pricing.[16]

---

[13]This does not require a rate system that charges real-time prices to all customers, a policy that could be both expensive and politically difficult. Substantial progress would be made even if only the largest customers received real-time prices, and if customers and providers were rewarded for *selling back* high-value consumption avoided in peak periods.

[14]Edison Electric Institute, *Statistical Yearbook of the Electric Utility Industry, 1999 Edition* (Washington, DC 1999)

[15]Hirst and Kirby, *Retail-Customer Participation in Bulk-Power Markets* (Oak Ridge, November 2000) at 30-31. Without naming utilities, the authors give some examples: "We know of one utility with about 500,000 customers, for which the largest 600 customers (all of which have 5-minute interval meters) account for 60% of total load. We know of another utility, for which the largest 15 customers represent 20% of total load." (at 31). The Hirst and Kirby paper provides the reader with an excellent overview of this issue, supported by several useful examples.

[16]Ibid.

Whether large customers bid their load directly into wholesale markets, or whether they are represented by franchise utilities or retail aggregators, bidding rules on the wholesale trading floor must be organized to reveal the customers' demand curve. The first step in this process is to require customers or their load-serving entities (LSE'S) to place binding bids into the market under the same general conditions as generators placing supply-side bids. Bidding rules should permit load to bid at multiple price points, stating how consumption will vary with different market prices.[17] Thus, when the market managers clear the market and set prices, both demand and supply bids will be involved in setting price and quantity in the settlement.

2. <u>Price-responsive load and demand-resales</u>: Power market design should accommodate price-responsive load and demand resales, through market rules that give willing customers and their suppliers reasonable opportunity to adjust consumption in response to market conditions. FERC should require adoption of multi-settlement systems or similar market rules to meet this need. FERC should also require RTOs to adopt and invest in standard real-time information systems and communications protocols that may be used by innovative energy service providers and load-serving entities to manage demand-side resources in response to market conditions.

A second needed reform in wholesale markets builds upon demand-side bidding, and extends the potential of demand responsiveness by recognizing the differences between projected market conditions and real-time events. Bidding rules should permit customers, generators, and reliability managers to plan consumption and generation decisions in advance. They should also permit additional adjustments to those plans in response to real-time conditions, such as changes in the weather, unplanned outages, changed consumer needs, or unanticipated price changes in the market. This is one of the principal advantages of "two-settlement" or "multi-settlement"[18] systems, discussed below.

In multi-settlement systems the market is "settled," or cleared, more than once, generally through the following steps:

(1) First, in the "day-ahead" market, bids are taken both for loads and for supply resources.
(2) Using settlement software to rank both demand and supply bids at various price-points, the market manager clears the market at prices and quantities that are physically achievable.
(3) At the time of this first settlement (usually a day ahead), accepted bids are not merely hypothetical – they are firm financial commitments to buy and sell power at the market clearing price. In a financial sense, power is bought and sold in this settlement.

[17] Across narrow price bands, load may be a price-taker, but be willing to curtail consumption at high prices. Bids with different price-points will reveal whether this is the case, and the markets settlements software can recognize their effects when settling the market. For example, current bidding rules in the PJM energy market permit either load or supply to place bids with up to 10 specific price-points set out.
[18] The terminology used varies by region. PJM has implemented the system described here, which is called the "Two-Settlements" system. The NE-ISO is developing a similar system, which it terms a "Multi-Settlements System."

(4) Following the initial settlement, and up to a cut-point in the "day-of" market, buyers and sellers can seek to modify their commitments in a second settlement. Any adjustments made in this settlement are also financially binding.
(5) Discipline is imposed on bidders in these settlements by requirements that generation and purchases conform to the obligations of their bids. Any deviations from the settlements are presumed to be met by purchases from the spot market and are charged to suppliers and customers at spot market rates.

Multi-settlement systems can add both price stability and flexibility to electric power markets as compared with a single, real-time settlement, such as the market used to date in New England. The market for energy services operated by ISO-New England since May 1999 has depended heavily on a single after-the-fact settlement, determined only after resources have been dispatched.[19] Day-ahead prices are forecast, but with fairly low confidence.[20] What has resulted is a structure in which neither supply nor demand-side resources have much opportunity to plan for and respond to volatile prices. By impairing the ability of demand-side resources to plan for load reductions, and to profit from re-selling demand at times of high prices, single-settlement systems reduce the ability of demand response to supply stability and reliability to the system.

In contrast, multi-settlement systems provide clear price signals to both suppliers and load in advance of physical generation and consumption activity. The first market in multi-settlement markets performs a hedging function for ultimate consumers and suppliers. (In effect, it reduces the exposure of load serving entities or retail customers to unexpected shortages in the real-time markets.) It also has the effect of reducing the potential windfall profits flowing to operating generators from unscheduled outages in other units. Used in conjunction with demand-side bidding, multi-settlements can also provide strong incentives to meet the supply resource commitments made in the day-ahead settlement.[21]

Profitable demand response in a multi-settlement system:

Multi-settlement markets support demand-side responses that can moderate the reliability problems and price spikes associated with thin operating margins. This occurs in at least two ways. First, if prices clear at very high levels in the day-ahead market, LSEs and their customers know in advance that it will be in their interest to reduce consumption and sell the released power back into the wholesale market in the "day-of" settlement. The same is true if the day-ahead market clears at normal prices, but prices spike the next day due to hot weather or unplanned outages. Anyone who purchased supply in the day-ahead market now has a clear

[19]Peter Cramton and Robert Wilson (Market Design, Inc.), "A Review of ISO New England's Proposed Market Rules," Executive Summary at 2.
[20]For example, In June of 2000, NE-ISO day-ahead **load** forecasts deviated from actual hourly loads by approximately 3.4% (roughly 440 MW). In contrast, hourly forecasted **prices** deviated from actual settlement prices by approximately 20% (or $8/MWH relative to the $39/MWh price for the month).
[21]Properly structured, it reduces the secondary rent or market price premium that will inure to the benefit of other resources in a single market clearing price from actions to reschedule units (or to declare a scheduled unit unavailable) in the real-time market. It may also strengthen the effective penalties from such rescheduling. (The unit that is declared out of service must pay the market clearing price for the resource that was no longer available.)

opportunity to profit by reducing consumption and selling back their power purchases into the spot market.

Importantly, demand-side sale-backs of this type provide a virtually automatic profit incentive to load-side managers to reduce consumption at times of high peak load. And, because those sale-backs are re-sales of power actually purchased in the day-ahead market, in a settlement that satisfied the system's physical constraints, it answers the frequently-raised concern that purchases of demand reductions may merely be paying for reductions in "phantom load." Conversely, an LSE that finds its consumption exceeding its day-ahead purchases in this case will be obliged to pay the high spot prices for the deviation. Thus, LSEs have an incentive to predict their loads accurately, and to control their loads in times of thin margins and high prices.

***Effects on reliability*:**

As can be seen from the examples above, multi-settlements markets create important opportunities for demand-side as well as supply-side managers to meet the needs of the electric system during times of high prices. Such a system is both financially and physically more stable than a system that attempts to balance the market through ramping up and ramping down supply alone.

Mitigating market power and market "gaming":

Market power monitors have raised concerns that a single settlement approach increases supplier incentives to manipulate markets through the physical or economic withholding of assets (e.g., by declaring units unavailable in the short term market).[22] Well-functioning futures markets (including day-ahead markets) can reduce supplier incentives for gaming of this sort in the real-time market. In multi-settlement systems, units that suddenly become unavailable must purchase supply in the real-time market to satisfy their output commitments; meanwhile, all other units that were cleared in the day-ahead market receive only their day-ahead prices, not the temporarily high spot prices. And demand-side resources can enter the market through sale-backs when prices rise, moderating the impact of temporary shortages. Under these conditions, the benefits to suppliers from short-term strategic withholding are greatly reduced.

Conclusion - Multi-settlements and Demand Resales:

Effective advance markets can serve to help reduce the financial exposure of load serving entities (and their customers) as well as suppliers, to variations in real-time clearing prices. Forward markets can also help reduce the financial incentives for suppliers to manipulate short-term market clearing prices.[23] They will also improve the strength and timing of price signals sent to

[22]In some electricity markets, there may be a strong incentive for the physical withdrawal of generation resources when reliability is most threatened. For example, in the California and New England markets, only a few firms may set the clearing price when resources are limited (e.g., peak periods). The incentives for such gaming may be very strong in constrained load pockets. The combination of only a few suppliers and a single clearing price may create a strong incentive to withdraw or declare units "unavailable" when such an event can lead to a significant increase in the market clearing price that all producers will receive.

[23]Arguably, it may simply shift the incentives for physical withdrawal of resources from the final market to the

end users to reduce loads when prices are high in real-time markets. Demand-side bidding and the multi-settlement process complement each other in each of these valuable functions. For these reasons, demand-side bidding and multi-settlement markets are important techniques to mitigate the reliability challenges, prices spikes, and market power problems seen in wholesale power markets in the U.S. in recent years.

3. <u>Reforming Load Profiles:</u> Where wholesale power costs are assigned to load-serving entities based upon customer class load profiles, FERC should require market rules that make alternative load profiles available for major load subclasses based upon energy use patterns and load control options.

An essential step in promoting meaningful demand-side bidding is the reform of the system of load profiles used by wholesale markets and wires companies to assign load responsibility among LSEs serving customers whose consumption is not measured by interval metering. Since the overwhelming majority of customers are not served by interval metering, the allocation of total sales to an LSE's customers has always been done through the use of load profiles -- assignments of a customer's monthly consumption among each hour of the month based upon assumptions about usage patterns among average customers in a particular class.

The widespread use of load profiles has a powerful dampening effect on demand-side responses to varying wholesale prices. Customers who modify their consumption in consideration of high peak-hour costs will never see the benefits of the change in their bills, unless the load profile that assigns their consumption to peak and off-peak hours also changes. Equally important, the LSE that serves such customers will never see the economic benefits of those shifts either. In these circumstances, neither the customer nor the LSE has the necessary incentive to manage load in response to changing market conditions, and the benefits of demand-side bidding are undermined for most customers.

While it is theoretically possible to require real time meters for all customers, advanced metering systems are still expensive and comprehensive deployment would take years to accomplish.[24] The simplest way to address the problem may be to require the use of multiple load profiles. In the residential class, for example, we know that two most readily controllable loads are water heating and air conditioning. At a minimum, regulators should require creation of alternative load profiles for customers with those loads controlled. With alternative load profiles in place, LSEs and others would have an incentive not only to search out customers who have these load management opportunities but to actively invest in equipment necessary to move customers from

---

periods that are used for determining price levels in the advance or hedging markets. Gaming the advance markets, however, will present more modest reliability problems. The advantage of time and financial signals sent by these earlier advance-markets should increase the pool of available resources to meet service needs when the physical resources are needed in the final contingency market.

[24]There are about 110 million residential electric meters in the U.S. To date, there is only one large-scale system in place to measure residential consumption on a real-time basis. Puget Sound Energy has deployed 300,000 advanced meters to residential customers, who are billed on a simple peak/off peak tariff that varies the price by 15% above and 15% below the normal average residential rate. The program began in May 2001, so experience is still limited; however, it appears to have modestly affected residential consumption patterns without dramatic changes in prices to customers.

one load profile to another.

Dealing with the problems created by broadly-averaged load profiles requires action by both state and federal regulators. Since load profiles are used by wholesale power markets to assign power costs among *wholesale* market participants, FERC has jurisdiction and the responsibility to ensure that those profiles promote equitable and efficient markets. This should be done as part of FERC's review of wholesale market rules. It is most critical that the wholesale market signal reaches the LSE. Relationships between LSEs and their customers are state-jurisdictional, and thus state regulation will govern the manner in which costs assigned to LSEs are passed on to, or allocated among, retail end-users. State regulators should adopt market rules (in the case of retail competition customers) or rate designs (in the case of default or franchise customers) that would promote equitable rates and the efficient use of energy, and the effect of this reform will be greater if customers see the benefit of lower-cost load profiles in their rates. However, it is not necessary to pass the RTO's customized load profiles through to end-use customers in order to give the LSE the proper incentive to manage customer load efficiently.

4. <u>Ancillary Services Markets</u>: FERC should require that market rules allow demand-side resources to supply ancillary services on an equal basis with supply-side resources. The criteria for supplying ancillary services should be written in technology-neutral terms, and should not require costly telemetry and metering for individual demand-side resources whose performance could be verified on a statistical, aggregated basis.

Electricity is a unique service in that production and consumption must be matched essentially instantaneously. Reliability of the power system is maintained by actively controlling some resources to continuously balance aggregate production and consumption. Historically this control was exercised only over large generators. Loads were most often free to consume electricity on their own schedules to meet their needs, while generation, under the control of the system operator, responded to the changing requirements imposed by loads.

However, from the perspective of the system as a whole, dispatchable load can provide many balancing services just as well as dispatchable or controllable generation. And as wholesale markets evolve to provide competition among generators, new opportunities can emerge for demand-side resources to participate actively in providing reliability resources to the power markets.

While FERC has recognized that competition will be desirable in setting the market values for different reliability services, those competitive reliability markets tend to be built on the same weak foundation as the market for wholesale generation generally (that is, lacking a demand-side element). Because reliability has traditionally been viewed as a resource that, in periods of thinning reserve margins, could only be satisfied by bringing on more supply resources, the reliability potential of the demand-side has often been overlooked. Resource adequacy was considered only a matter or bringing forward robust administrative mechanisms or, in the emerging market world, adequate profit incentives to promote new generation so as to meet system margin requirements.

Individual ancillary services

Reliability from Distributed Resources

Whenever a system operator calls for the deployment of contingency reserves there is always some chance that the resource that is supposed to supply the reserve will fail to do so. The small size of individual distributed resources reduces the consequence of this problem and makes them a more reliable source of contingency reserves. Take, for example, the case of a system operator purchasing 100 MW of supplemental operating reserve from a 100 MW fast-start combustion turbine. This turbine might start within the required time on 90% of its attempts. In one case in ten the system operator is 100 MW short. It does the system operator little good to reduce its expectations to 90 MW, although that is the average response.

A collection of 12,500 10-kW distributed resources that individually have only an 80% chance of responding each time makes a better aggregated resource. In this case, 20% of the individuals fail to respond but the system operator still sees the full 100 MW response each time.

As a practical matter, ancillary services can be provided by supply-side generation, distributed generation, or customer-side load management arrangements, depending upon particular reliability needs and resource characteristics. These services are required to maintain bulk power system reliability and are being opened to competitive markets in regions where RTOs operate.[25] The most important opportunity to provide ancillary services on the demand side has been termed "dispatchable load" -- load that can be interrupted or reduced reliably to balance the system as load and supply conditions change. Interruptible customers and storage devices may best be able to provide Load Following and Supplemental Reserve services. However, they are not likely to provide Reactive Supply and Voltage Control From Generation to the bulk power system. Network Stability is a service that both distributed generators and storage devices should excel at if they are connected to the power system through an inverter and are in the correct physical location. Blackstart appears to be a service that small distributed generators may be qualified to sell, but which cannot be provided by customer-side load management alone.

Five of these services (Regulation, Load Following, Frequency Responsive Spinning Reserve, Supplemental Reserve, and Backup Supply) deal with maintaining or restoring the real-energy balance between generators and loads. These services are characterized by their response time and response duration, and by the communications and controls between the system operator and the resource needed to provide the service. Because regulation requires continuous (minute to minute) adjustment of real-power transfers between the resource and the system, loads may not be very good at providing this service. Load Following, however, could be provided directly or through the use of a spot market price response on a time frame of less than an hour, consistent with FERC's requirements that RTOs operate real-time balancing markets. The contingency reserves are especially amenable to being provided by distributed resources, including load

[25] FERC started this process by requiring the separation of six ancillary services from transmission in its Order 888; the Commission later expanded that process with its Order 2000 on regional transmission organizations (RTO'S).

management programs.

Some important considerations

While the benefits of dispatchable load as a reliability service are conceptually straightforward, it will require careful thought to create market rules that will permit demand-side resources to compete fairly in ancillary services markets. To begin with, system operators must articulate the requirements for reliability services in technology-neutral language. That is, the required performance must be specified clearly enough that separate commercial entities with either demand or supply resources can agree on what will be provided, and at what price. The requirements must specify performance rather than the methods to yield desired outputs. For example, a system operator should request “100 MW of response that can be delivered within 10 minutes” rather than “100 MW of unloaded, on-line capacity from a large fuel-burning generator.”

In addition, system rules should respect the practical realities of distributed, demand-side resources, and not require the types of communications and controls traditionally used to control central station generators. Those who provide demand-side ancillary services should be permitted to demonstrate the availability and performance of aggregated, distributed resources through reliable statistical means, without the necessity of hooking up each customer to real-time meters and direct links to system control room operators. It is not necessary to know exactly how each customer site is responding to system requests in order to rely on demand-side resources to help balance the system.[26] Many loads, such as commercial or residential air conditioning systems, could be monitored through statistically-sound sampling protocols that could provide highly reliable information to control room operators as well as market managers. Indeed, in some respects, distributed resources may be more reliable than traditional supply-side units, due to the diversity benefits of multiple sources of non-traditional supply. (See Box -- Reliability From Distributed Resources, above).

Conclusion - Ancillary Services:

If ancillary markets are established so that demand-side resources can participate actively, load management resources benefit because they receive revenue from the sale of valuable ancillary services. The power system also benefits in several ways. FERC ordered the unbundling of ancillary services from transmission to promote competitive markets, which should improve economic efficiency and lower electricity prices. These markets should be open to any technology capable of providing the service, not just to traditional generators. This will expand supplies and reduce the horizontal market power of generators. Finally, because ancillary services consume generating capacity, providing reserves through demand-side resources also improves overall resource utilization and reduces the excess fuel use and pollution associated with spinning reserves that are not serving load. When loads provide these reserves, generating

[26]This is not to argue that demand-side resources should be held to lower standards than supply-side providers of ancillary services. For example, for a generator to supply contingency reserves, it must have capacity available to respond to the contingency; the generator cannot be operating at full load. Similarly, a load manager selling contingency reserves must have capacity it can make available when the contingency occurs-- in this case by ensuring that there is active load in its portfolio that can actually be curtailed.

capacity is freed up to generate electricity.

5. <u>The Efficient Reliability Standard</u>: RTOs, reliability managers, and transmission owners often seek cost recovery in FERC-approved tariffs for investments intended to enhance system reliability. Before granting recovery that would broadly "socialize" those costs, FERC should require applicants to show that the benefits are broadly dispersed, and that they have selected the lowest-cost resource, including demand-side resources, reasonably available to meet the need in question.

Resource adequacy and system reliability across electric networks are classic public goods, provided to all interconnected users on essentially the same basis, and not easily withheld from any interconnected user. Efficiently constructed wholesale electricity markets, including adequate demand-side bidding systems, can moderate both the volatility of markets and the degree to which reliability managers must intervene in the market to ensure reliable service. Nevertheless, reliability and power market managers often find it necessary to take administrative actions to promote reliability. And typically, they seek to recover the costs of these administrative actions in broad-based rates charged to all users of the grid. These administrative actions take many forms:

- Requiring the provision of specified ancillary services by market participants by rule; and/or purchasing them on behalf of all market participants (and then imposing a tariff to pay for them);
- Socializing congestion costs, supported through uplift charges, so that customers in load pockets do not pay higher prices for power behind a constrained interface;[27]
- Entering the market directly through an RFP for the provision of reliability services, such as the emergency generators and dispatchable load contracts sought to be deployed in several power pools in recent summers;[28]
- Identifying needed transmission links and supporting their construction through broad-based transmission tariffs or other forms of "uplift" assigned to users throughout the pool;[29]
- There are many other variations on this theme.

System operators have traditionally focused on supply-side resources in meeting reliability

---

[27] This has been the practice in New England for many years.

[28] For example, in the summers of 1999 and 2000 the New England and California ISOs proposed collecting pool-wide uplift charges to bring in and operate emergency generators on barges anchored in the Connecticut River and San Francisco Bay. Several pools have launched programs to acquire demand interruptions from customers who will agree to load controls operated directly from the ISO.

[29] In 2000, the New England ISO accepted a recommendation to support the construction of several transmission upgrades throughout the region, as "Pool Transmission Facilities" because they would relieve transmission congestion in certain areas, and improve the resilience of the transmission system. In NE-ISO parlance, the cost of these upgrades will be "socialized" -- that is spread among all users of the regional transmission system through a regional "uplift" charge. More than $120 million in capital costs will be raised, under a NE-ISO tariff, for this purpose. The New England 2001 Regional Transmission Expansion Plan (Draft 8/24/01) is now considering adding transmission enhancements to support the reliability of the region's electric system.

requirements for electric networks, especially in periods of stress. However, for many system needs, there is a demand-side corollary that could perform that same service at lower cost, provided that market rules were defined to include such resources, and broad-based funding were made available to support them on the same basis as the more traditional solutions.
Energy efficiency, load management, demand-side bidding, and distributed resources – in addition to traditional generation and transmission resources -- are all potentially cost-effective means of meeting reliability needs identified by system operators and power pool managers.

So long as vertically integrated utilities were basing their investment decisions on the principles of integrated resource planning, many reliability-enhancing decisions were governed by least-cost decision-making and associated regulatory review. With the breakup of the franchise, the demise of IRP, and the assumption of new responsibilities by RTOs and other regional organizations, there are now numerous occasions where broadly-funded interventions may be taken without serious consideration of less expensive and more reliable alternatives based on distributed resources and demand-side alternatives.

For this reason, reliability rules and investment decisions that will, by administrative action, impose costs on consumers and other market participants, should first be tested by the following standard for the efficient provision of reliability (See Box, "The Efficient Reliability Standard"):

The Efficient Reliability Standard

Before "socializing" the costs of a proposed reliability-enhancing investment through tariff, uplift, or other cost-sharing requirement, FERC should require the applicant to demonstrate that:

t --the relevant market is fully open to demand-side as well as supply-side resources;

t --the proposed investment or standard is the lowest cost, reasonably-available means to correct a remaining market failure; and

--benefits from the investment or standard will be widespread, and thus appropriate for support through broad-based funding.

If this standard were adopted as a screening tool by FERC and the nation's RTOs when considering proposed reliability-enhancing rules and investments, it would provide a much-needed discipline in situations where expensive wires and turbines solutions are proposed to address reliability problems, and more robust, less expensive, distributed solutions are overlooked.

A standard of this sort has been under discussion in at least one of the nation's ISOs. The Advisory Council to the Board of the New England ISO addressed this issue, concluding that it should be the responsibility of the ISO, as New England's RTO, to analyze feasible alternatives to transmission investments for reliability, and to present them as part of a regional reliability planning process. "The ISO should develop a regional plan for transmission, with updates and modifications as market participants provide market-driven proposals to relieve reliability and economic congestion conditions; this plan should identify high-value energy services and reliability needs that should be addressed, and which could be remedied through either a transmission, generation, distributed-generation, or demand reduction approach."[30]

In addition, the Advisory Council concluded that the ISO should analyze alternatives to transmission, and make those analyses available in the public decision-making process on socialization proposals: "In situations where the ISO-NE is called upon to analyze transmission proposals whose costs are proposed to be socialized (that is, recovered through a broad-based mechanism paid for by captive ratepayers, as opposed to proposals to recover merchant transmission investments through market mechanisms), the ISO-NE should have a heightened duty to make sure that its analyses consider the impacts of the proposal and a broad array of practical transmission and non-transmission alternatives to it. These analyses should be designed to analyze whether the proposed transmission facility is needed and least-cost."[31]

Opportunities to adopt and apply this principle arise in numerous circumstances. At the legislative level, Congress and state legislatures have seen many proposals to amend underlying enabling legislation relating to reliability. Statutory revisions in this arena should adhere to the principle that demand-side and supply-side reliability options should be treated equally in considering how best to address reliability needs.

FERC should also take the initiative on this point. Numerous state PUCs have long understood that least-cost principles should govern utility decisions to make investment decisions that they plan to recover from ratepayers. Increasingly, those decisions are being made today by RTOs, ISOs, Transcos and wholesale power pools, subject to FERC jurisdiction. FERC should require RTOs to ensure that decisions to socialize reliability improvements have been disciplined by a hard look at traditional, distributed, and demand-side alternatives.

Finally, it will be important that reliability and transmission planning processes be fluid enough so that RTO analyses under the Efficient Reliability Standard are updated and revised over time. Transmission planning often takes a long time, while market-driven, economically attractive alternatives may have shorter lead times, and may appear after a "build" decision is reached on the transmission alternative. To support both competitive markets and reliability objectives, the RTO transmission planning process must allow for changes in conditions that may reveal different reliability solutions (whether demand-side resources, distributed generation, or something else). In such cases, FERC's standards for cost recovery must encourage the "later look" and allow for cost recovery of planning and development costs when a project is prudently curtailed in favor of a less costly alternative.

6. <u>Transmission-Level Congestion Pricing</u>: Transmission prices that hide from customers the costs of congestion, and the value of congestion relief, diminish the reliability contribution that could be made cost-effectively by load

[30] ISO-NE Advisory Committee, Comments to the ISO – New England Board Regarding the Design of a Regional Transmission Organization for New England, October 11, 2000. (Notably, the ISO-NE is currently undertaking a planning process for Southeast Connecticut in a way that does not appear to include a robust consideration of either traditional generation, distributed generation, or demand-side alternatives to transmission solutions.)

[31] Ibid.

management, efficiency, and generation in load pockets. FERC should promote transmission rate designs that reveal the cost of congestion across different times and locations.

Transmission congestion refers to the situation in which it is not possible to complete all the proposed transactions to move power from one location to another on the grid. Such commercial-transaction restrictions can arise because of thermal, voltage, or stability limits on transmission elements. Congestion may occur even when the actual flow on a line is well below the line's capacity, whenever security-constraints in the larger control area require modification of the economic dispatch order. This situation occurs most frequently as the result of contingency analysis rather than because of steady-state line flows. The generation dispatch is modified because a line *will* overload *if* a specific contingency occurs (e.g., a generator or transmission line trips). Because there is often no time to take corrective action to prevent cascading failures if such a contingency occurs, it is necessary to preemptively modify the generation dispatch ahead of (and in anticipation of the potential) contingency. It is this off-economic dispatch that results in locational price differences.

Transmission Resource Constraints are Not Fully Recognized in Wholesale Prices

As wholesale markets have restructured, and as load has grown, transmission congestion has become an increasing problem. The traditional vertically integrated utilities accounted for transmission constraints when they made their daily operating (unit-commitment) plans, as well as operational dispatch decisions. Thus, they used their generating resources in ways that would not overload the network. However, in today's increasingly competitive environment, suppliers schedule resources without a detailed knowledge of, or interest in, transmission constraints in the larger region.

Transmission constraints impose significant costs on electric systems. However, in most places, so far at least, those costs have not been reflected in either wholesale or retail prices. They have typically been collected through a system uplift charge assessed on all buyers. This tradition must be changed in order to expose cost-effective solutions to congestion problems. The variability of wholesale costs caused by such constraints needs to first be recognized in wholesale prices. Financial congestion rights can assist transmission planners and potential generators looking for promising locations for new generation sources. Location specific pricing of energy services in the face of such constraints may provide the necessary incentives to LSEs and final consumers to manage loads during periods when transmission lines constrain access to the broader market.

In the long term, construction of new generators and transmission lines, or the addition of load reduction and load management programs, can reduce congestion. In the short term, system operators can treat congestion in three possible ways: they can mandate engineering solutions; they can socialize the costs of congestion; and/or they can use prices to let suppliers and consumers decide where to make investments and which transactions to forego.

The simplest (engineering) approach is to ignore congestion in setting energy prices (i.e., assume that all proposed transactions can be completed as if the amount of transmission capacity was infinite).[32] If proposed transactions threaten to overload transmission lines, the security coordinator implements NERC's transmission loading relief (TLR) procedure. This procedure adopts an engineering approach to congestion relief. Transactions that contribute 5% or more to the congestion are cut. Many market participants oppose TLR because they believe that the incumbent utilities manipulate the TLR calls and implementation to favor their own transactions. In addition, FERC opposes the current TLR procedure because it is economically inefficient.

An alternative approach is to socialize congestion costs. With this approach, the system operator pays generators on either side of the constraint to increase output (constrained on) or decrease output (constrained off) to relieve the congestion. The system operator pays these constrained-on generators for any opportunity or out-of-pocket costs associated with this uneconomic dispatch. The costs so incurred are then allocated to all transmission customers through an uplift charge. Although simple to implement, this approach is economically inefficient because it fails to send price signals to transmission users (LSEs and, ultimately, their customers) on the true costs associated with their transactions. The absence of location-specific prices also robs investors of important information on where to locate new generators or new demand management programs, and what transmission projects to build.

Finally, a more economically efficient way to deal with congestion is to use locational prices that provide signals to transmission users on the actual costs of transmission service. Because transmission constraints vary with time, and vary from location to location, pricing systems that value access to different transmission paths at different times can be complex.[33] The details of the method chosen in a region are less important than the basic policy conclusion: locational transmission pricing reveals the costs of congestion, and thus reveals the potential value of demand-side measures that manage load and provide congestion relief.

Locational Pricing reveals the cost of congestion and the value of congestion relief

The preceding discussion of transmission systems and costs touches on two important features relevant to the role of demand resources: (a) Congestion on the transmission network raises very important reliability problems, not just for the load centers directly affected, but potentially for customers across the entire affected network; and (b) Congestion on the transmission grid is not even across the network, and it varies with time. For these reasons, energy efficiency and load management resources may have great value when they reduce load in the particular locations and at the particular times that congestion problems would otherwise arise.

The application of locational pricing is an important step in the development of competitive electricity markets. When congestion costs are assigned to the responsible load, a more accurate price signal is received within the load pocket. Thus, cost-effective means to reduce congestion

[32] This may require proposing transmission engineering enhancements later, when chronic transmission constraints occur (so that the transmission infrastructure can be restored to the unconstrained level of service assumed in the first instance.)

[33] For example, locational prices can be set on a zonal basis (as in California) or on a nodal basis (as in PJM). New York and New England plan to use a mixed system, in which generators will face nodal prices and consumers will face zonal prices.

will have the opportunity to compete to reduce the congestion and improve reliability.[34] Generation, transmission, and load management options will all have the incentive and the opportunity to offer cheaper solutions to customers and load-serving entities within the load pocket. Because locational pricing sets an appropriate "price-to-beat" benchmark, replacing a system in which congestion costs are not revealed to customers, efficiency and load management investments can compete on a fair basis with transmission and generation options to provide reliability services in the load center.

> **7. <u>Regional Reliability Charges</u>: FERC should encourage RTOs to examine region-wide, reliability-enhancing investments in demand-side resources that would improve reliability and lower power costs. When supported by cost-effectiveness analysis, RTOs should be permitted to recover those investments on the same basis as regional transmission investments, ancillary service costs, or other RTO expenses**.

Historically, utility energy efficiency programs were administered at the franchise level. More recently, many states have recognized the value of statewide programs and funding sources. But there is no essential reason to draw the boundaries for efficiency programs at the state line. Power markets today are regional; transmission grids and system operations are regional; and reliability rules and the costs of reliability programs are imposed across franchise and state boundaries. Moreover, the markets and delivery channels for many end-use technologies (such as industrial motors, chillers, and household appliances) are regional in nature. Finally, the economic, environmental, and reliability *benefits* of improved electric efficiency flow to consumers across power pools and transmission grids, and do not stop at the state boundary. For all of these reasons, policy-makers should consider the merits of broad-based, regional funding for efficiency programs that will benefit regional power markets, regional reliability, and regional transmission systems.

The benefits of regional energy efficiency programs have been recognized by energy professionals and decision-makers in a variety of contexts. The leading example has been the Northwest Power Planning Council, which has sponsored very significant programs throughout the multi-state region served by the Bonneville Power Administration. In recent years, multi-state programs in the BPA region have also been developed and funded through the Northwest Energy Efficiency Alliance, a non-profit corporation governed by a board of utility, government, and other stakeholder representatives. As of 1999 it had an annual budget of $22 million per year.[35] Regional efficiency organizations have also been established in the Northeast and the Midwest.[36]

---

[34] Particularly if the Efficient Reliability Standard (discussed above) is applied to proposals that would socialize congestion relief and mute the signals sent by locational pricing. Locational pricing and the Efficient Reliability Standard work together to advance the most reliable and lowest cost solutions to congestion problems.

[35] Raab and Peters, "A Comparative Study of the Northwest Energy Efficiency Alliance and the Northeast Energy Efficiency Partnership," (NARUC 1998) at p.13.

[36] The Northeast Energy Efficiency Partnership sponsors about $20 million per year in efficiency programs in the region stretching from Maine to Maryland. It is funded principally by utility contributions, but receives some federal and state agency support as well. Ibid. The youngest of these organizations is the Midwest

While the current regional efficiency partnerships offer promise, they are both voluntary and relatively small. What is lacking is a consistent funding mechanism to support delivery of demand-side resources in regional wholesale markets. Some modest load-management resources are now supported at the ISO level (see, e.g., the load-response programs sponsored by the New England, New York, and California ISOs in the summers of 2000 and 2001), but longer-term energy efficiency investments have been left to individual utilities and to state system benefit funds.

This is a missed opportunity for two reasons. First, regional power managers -- RTOs, ISOs, Transcos, and reliability organizations -- are engaged in the process of securing generation, ancillary services, reserves, and transmission projects on a regional basis. Where efficiency investments would meet those system needs at lower cost, the failure to invest in efficiency is driving up the *cost* of regional collection mechanisms, and of reliable power for the region. Second, efficiency investments can provide benefits to consumers across a region by lowering the *price* of power in regional power markets. Evidence from existing regional markets in California, PJM, and New England supports the conclusion that modest regional wires charges supporting regional efficiency programs could be highly cost-effective.[37] Yet, in the absence of a regional, non-bypassable collection mechanism, individual utilities and states will continue to benefit from their neighbors' programs, whether or not they support equivalent programs of their own.

Wholesale markets could be designed to capture large consumer savings through broad-based market transformation or energy efficiency programs without much difficulty. With so much money to be saved and so many reliability benefits to be achieved these questions should be high priority issues for FERC and state regulators.

8.<u>Coordinating Retail and Wholesale Market Rules</u>: The reliability and economic benefits of demand-side resources will be best developed where wholesale market rules, and retail pricing and demand-bidding rules are consistent. FERC should encourage efforts within each RTO to bring state and federal regulators and interested stakeholders together to identify unintended conflicts and to develop a coherent set of market rules, tariffs, and policies in support of cost-effective distributed and demand-side resources.

Wholesale market structure, ancillary services, and transmission pricing policies are within the province of FERC, while retail market structure, default service, and retail pricing policies are within the province of state regulation. Wholesale and retail electricity markets are not wholly different realms, but are connected links in an economic chain connecting customers, retailers, wires companies, and generators. In the areas of demand management, customer price response, and energy efficiency, there are numerous instances in which policies and practices in one of

---

Energy Efficiency Alliance, also founded in recognition of the regional nature of electricity markets and efficiency benefits.

[37]These points are discussed in the sections on demand-side bidding and the Efficient Reliability Standard, above.

these markets will directly affect the effectiveness of price signals or policies in the others market. For example:

- The degree of demand response to hourly prices in the wholesale market can be very seriously dampened by averaged rates, price caps and real-time metering costs at the retail level;
- transmission pricing policies can reinforce or undercut the value of demand management by retail customers;
- the benefits of more accurate price signals to customers can be undercut if load profiles assigning costs to LSEs at the wholesale level are not also made more accurate;
- to be effective, price-responsive load programs at the regional level must be coordinated with traditional retail load-interruption contracts;

.

Experience will reveal many more connections of this type.

This paper has focused on policies that FERC and RTOs can implement to advance the positive role of customer-based load response and energy efficiency investments in power systems and markets. These are important steps for FERC and RTOs to take, but they should also work with state regulators and others to advance coordinated approaches and bridge the gaps between federal and state jurisdiction. To this end, FERC should encourage efforts within each RTO to bring state and federal regulators, RTO managers, and interested stakeholders together to identify unintended conflicts and to develop a coherent set of market rules, tariffs, and policies in support of cost-effective distributed and demand-side resources. Such a coordinated approach will not, of course, intrude upon state jurisdiction. But it would give state regulators and retail providers the opportunity to recommend RTO and FERC actions that they would like to see, to consider governance structures that could promote coordinated solutions, and to identify tools to enhance markets and improve reliability, adapting the industry's hard-won principles of cost minimization to the realities of emerging power markets.