ACCOMMODATING DISTRIBUTED RESOURCES IN WHOLESALE MARKETS

September 2001

Frederick Weston

with assistance from:
Cheryl Harrington
David Moskovitz
Wayne Shirley
Richard Cowart
Richard Sedano

The Regulatory Assistance Project
177 Water Street, Gardiner, Maine 04345
Phone: 207-582-1135 Fax: 207-582-1176
50 State Street, Montpelier, VT 05602
Phone: 802-223-8199 Fax: 802-223-8172
www.raponline.org

PREFACE

This report benefitted from the thoughtful review of a number of people, including my colleagues at The Regulatory Assistance Project, members and staff of the National Association of Regulatory Utility Commissioners' Committee on Energy Resources and the Environment, Thomas Basso and Gary Nakarado of the National Renewable Energy Laboratory, and Joseph Galdo of the US Department of Energy. In particular, thanks go Richard Cowart, Eric Hirst, and Brendan Kirby, upon whose earlier work on markets, reliability, and demand response this paper hopes modestly to build.

Thanks to all,

Rick Weston

This paper is one of a series published by the Regulatory Assistance Project on Distributed Resource Policies for state and federal regulators. The reader is encouraged to read the others in this series which can found at RAP's website: www.raponline.org

This report was prepared by the Regulatory Assistance Project under contract with the National Renewable Energy Laboratory. The views and opinions expressed herein are strictly those of the authors and do not necessarily agree with, state, or reflect the positions of NREL or those who commented on the paper during its drafting.

TABLE OF CONTENTS

I. INTRODUCTION

Changes in electricity markets, in technology, economics, and regulatory structures, have created a new interest in small-scale generation and efficiency resources, dispersed throughout the lower-voltage networks. These "distributed" resources can provide cost-effective reliability and energy services, in many cases obviating the need for more expensive investments in wires and central generating stations. Given the unique features of distributed resources, the challenge facing policymakers today is how to structure wholesale markets for electricity and related services be structured so as to reveal the full value that they can provide to the system. Put another way, how can the markets be organized and managed so as to enable distributed resources to compete to the greatest extent possible?

This report offers answers to this question. It looks at the different functions that distributed resources ("DR") can perform and examines the barriers to them. It then identifies a series of policy and operational approaches to promoting DR in wholesale markets. Those remedies include:

- Demand-side bidding and multi-settlements;
- Demand response (participation of load management in spot markets);
- Opening the ancillary services market to DR;
- Resource aggregation and management;
- Increasing market liquidity;
- More economically efficient transmission and distribution rate design; and
- Public benefits programs, including funding mechanisms, in support of investment in long-term end-use energy efficiency.

This is less a menu of choices than it is a set of integrated strategies to improve economic efficiency, lower total costs, and enhance reliability of the nation's electric system. Each provides value in its own way. And, though the focus of this paper is on competitive wholesale markets, it will become apparent that many of these recommendations also are applicable to vertically integrated monopoly structures (which, it is worth noting, still constitute the vast majority of the country's electric sector).

II. WHAT IS DR AND WHAT KINDS OF SERVICES CAN IT PROVIDE?

A. Distributed Resources

Distributed resources, or DR, describes the broad set of electricity-generating and electricity-saving measures that are located near or on customer premises – that is, are *distributed* throughout the network, close to loads. Distributed resources include smaller-scale generation, combined heat and power, energy storage, load management, and energy efficiency. There is no established measure for the size of distributed resources: typically they are thought to include technologies of up to 10 MW, but some customer-owned generation is significantly larger, for example, 100 MW. DR can be owned by a customer (load), a utility, or a third party (*i.e.*, independent power producer). Efficiency and load management resources of course are "found" on a customer's premises; generation and storage resources, however, can be located at customer's facilities, utility substations, or elsewhere on the lower-voltage system.

It is its size and proximity to customer loads that distinguishes DR from traditional central generation and delivery. DR can deliver electricity directly to the consumer who owns it or to the distribution network, thus avoiding use of the transmission system. DR facilities are smaller than central stations, are capable of remote operation, and can serve a variety of uses.

Included in the broad category of distributed resources are a number of different technologies such as microturbines, reciprocating engines fueled by gasoline, diesel, or natural gas, fuel cells, gas turbines, photovoltaics, wind turbines, and the wide array of load management and end-use efficiency measures.

B. Uses Served By Distributed Resources

Although we tend to think of electricity as a single service and now, with the restructuring of markets across the country and around the world, more as a commodity, it in fact is made up of several components. The capacity to generate electricity and the energy actually produced are electricity's two chief components, but the ability of the system to produce and deliver that energy in a usable form (at the proper voltage, frequency, etc.) depends upon other actions, called ancillary services, being taken.

Although capacity and energy can be, and have historically often been, sold separately, new wholesale electricity markets typically sell primarily energy – kilowatthours – on an hourly basis. Changes in energy prices across time reflect the increasing value of capacity as peak demands (or other constraints) are approached and, over the longer term, if the market is functioning properly, efficient suppliers should recover their capacity costs through energy sales.[1] Ancillary services, which tended previously to be bundled with energy and capacity (or simply provided by the utilities themselves), are now separately purchased. This is in part because responsibility for network management has shifted to an independent operator, but also because there is an expectation that the new competitive markets, with their many participants, will create more efficient ways of providing these services.

Ancillary services are needed to meet bulk system reliability needs. There are at least nine of them, and they can be separated and individually purchased. Of them, eight can be served by distributed resources.[2] They are:

- *Reactive Supply and Voltage Control from Generation:* Injection and absorption of reactive power from generators to control transmission voltages;
- *Regulation:* Maintenance of the minute-to-minute generation/load balance to meet the North American Electric Reliability Council's (NERC's) Control Performance Standards 1 and 2;
- *Load Following:* Maintenance of the hour-to-hour generation/load balance;
- *Frequency Responsive Spinning Reserve:* An immediate (10-second) response to contingencies and frequency deviations;
- *Supplemental Reserve:* A response to restore generation/load balance within 10 minutes of a generation or transmission contingency;
- *Backup Supply:* A customer plan to restore system contingency reserves within 30 minutes if the customer's primary supply is disabled;
- *Network Stability:* Use of fast-response equipment to maintain a secure transmission system; and
- *System Blackstart:* The capability to start generation and restore all or a major portion of the power system to service without outside support after a total system collapse.

These services are required to maintain bulk power system reliability and are being opened to competitive markets in regions where RTOs operate. As a practical matter, not all of these services can be provided by all forms of distributed resources. Distributed generators, interruptible customers, and storage devices may best be able to provide Load Following and Supplemental Reserve services; depending on their size and location, they may not be able to sell Reactive Supply and Voltage Control From Generation to the bulk power system. Network Stability is a service at which both distributed generators and storage devices should excel if they are connected to the power system through an inverter and are in the right physical locations. Blackstart appears to be a service that small distributed

1. In New England, where there are both a market for energy and a market for installed capacity, the independent system operator has recently proposed the elimination of the capacity market.

2. This list does not exactly match FERC's own. System Control is not included because DR owners can not sell that service. System Blackstart, Backup Supply and Network stability are included because DR owners might be able to sell these services even if FERC does not explicitly recognize them. The services also do not precisely correspond to the current NERC Interconnected Operations Services, included among which is Frequency Responsive Reserve. The precise definitions are still in flux, although the concepts are well accepted.

generators may be qualified to sell since many such generators are inherently capable of operating independently of the power system. To be useful to the power system, however, the blackstart units have to be located where they will be able to re-start other generators. Some DR generators are not large enough or located in the right places to be useful in this regard. For those that are big enough and in the correct locations, however, this could be an excellent service to sell.

The five remaining services (Regulation, Load Following, Frequency Responsive Spinning Reserve, Supplemental Reserve, and Backup Supply) deal with maintaining or restoring the real-time balance between generators and loads. These services are characterized by response time, response duration, and communications and control 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 want to provide this service. Load following could be provided directly or through the use of a spot market price response on a time frame 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 management programs.

Similar restrictions apply to DR supplying ancillary services as apply to central generation stations supplying those same services. 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 DR selling contingency reserves must have capacity it can make available when the contingency occurs, either by increasing its power output or by temporarily curtailing load.

Finally, who actually owns a distributed resource can have implications for how it can be used. Utility-owned DR is under the control of the utility and can be dispatched directly. DR that is owned by customers, load-serving entities (LSEs), or other market participants may require additional contractual and operational mechanisms in order to facilitate dispatch.

C. The Value of Distributed Resources

Distributed resources can provide benefits to the electric system in a variety of ways:

- *Peak load management*. The costs of serving load are highest at times of system peak. Under the traditional model of economic dispatch of generation for regulated utilities, DR would be operated whenever its marginal costs were less than those of alternative resources. Any savings that accrued would reduce a company's total cost of service and, all else being equal, would lower rates. In competitive markets, DR will be dispatched whenever its bid prices are lower than competing alternatives, but now the savings, large to begin with, are even more substantial. The reason for this is that wholesale markets clear at the price of the marginal bid in an hour; all buyers pay the price of the last resource needed to clear the market in that period. Consequently, any reduction in loads or increase in available supply lowers prices for all.

- *Market liquidity.* Enabling more resources to participate in wholesale markets increases liquidity. Liquidity spurs innovation in and proliferation of service offerings, puts downward pressure on prices, and mitigates market power. In the way that wholesale prices decrease as loads decrease or as supply increases, the converse is also true: prices rise as load increases, particularly at times of peak. It is this phenomenon that can encourage suppliers to withhold particular resources from the market, thus drawing higher cost marginal supplies into the bidding. The overall price, which is paid to all suppliers dispatched in the period, increases as a consequence. This is an effective strategy if the incremental revenues one receives as a result of the incremental price increase exceed the net revenues foregone by withholding a particular unit or units. Such withholding is an exercise of market power. Distributed resources provide one check against its abuse.

- *Transmission and distribution cost savings*. Distributed resources, being located primarily on the distribution system can obviate the need for, or at least defer, new investment in transmission and distribution (T&D). These savings, particularly in constrained areas, can be substantial – enough by themselves to justify the DR investment, before any of the energy benefits are taken into account.

- *Ancillary services*. The flexibility and dispersed nature of distributed resources make them excellent providers of reliability services.

III. DISTRIBUTED RESOURCES AND WHOLESALE MARKETS

A. Wholesale Markets

Electric energy and capacity is traded at wholesale – that is, among utilities and other providers of service – across the nation. In those states and regions where utilities remain vertically integrated and fully regulated, trading generally takes the form of bilateral contracts for power transfers of varying durations – years, months, weeks, days, even hours. In competitive markets, the same sorts of trading also occur but, in addition, a variety of spot and futures markets for the various commodities have also been created.

The primary, or spot, markets for energy, capacity, and ancillary services are managed by the system operator. They are used to set clearing prices for resources to meet residual energy needs and reliability services on a day-ahead and day-of basis. Secondary markets – markets for futures and other hedging instruments – are emerging as liquidity (the number of market participants and service offerings) increases. National Westminster and Morgan Stanley have both created clearing houses for trading energy in future weeks, months, and years. Products are available in the New England, New York, and Pennsylvania-Jersey-Maryland (PJM) regions.

B. Information and Transactional Requirements

All operators of electric generating facilities need access to particular kinds of information – bidding rules, dispatch requirements, their own costs, operational requirements, etc. – if they are to participate successfully in the market. For large, central generating stations, the costs of obtaining and managing such information are small in relation to their total costs. For distributed resources – dispersed, relatively small, and often expensive to operate – information costs may be quite significant and pose a barrier to participation in the market. DR needs low-cost access to relevant information, or alternative operational approaches that reduce their need for such information. What is fundamentally at issue is how to make sure that distributed resources can see and exploit the full value of the benefits that they can provide the system.

Opinions vary on whether and to what degree distributed resources will participate meaningfully in the spot and futures markets for energy. The output of an individual distributed resource – the energy from only a few kilowatts or megawatts of capacity – may be too small to be tradable in large markets where many transactions occur and great quantities of electricity are bought and sold. That does not mean, of course, that the DR doesn't have value at least equal to the market price for power at any particular time; merely it suggests that the transaction costs associated with making its output available for anonymous purchase on the spot market overwhelm the cost-effectiveness of the sale.

This difficulty can be overcome through the aggregation of distributed resources under the control of a single entity. One such entity could, of course, be the system operator. Another could be the distribution local utility. And, as the markets expand, one would expect load-serving entities (LSEs) and other marketing firms also to become aggregators. Aggregators act as intermediaries between the wholesale markets and system operator, on the one hand, and retail consumers and owners of DR on the other. They can facilitate transactions, manage load, and re-sell power as market conditions dictate; in this way, they are analogous to arbitragers in financial markets. What becomes important, then, is that the market rules and regulatory policies be designed so as to promote their participation in the market. See Section III.C.1.d., below.

Aside from their participation in the wholesale markets for energy and ancillary services, distributed resources can provide significant benefits to the transmission and distribution system. Indeed, it may be in deferring or altogether avoiding new T&D – particularly distribution – investment that DR's greatest value lies.[3] In this instance, the information needs go to the costs of distribution, not to the prices in wholesale markets. Assuring rational planning and investment in distribution, along the lines of integrated resource planning, is a first step that regulators can take to promote least-cost outcomes.[4] In addition, innovation rate offerings targeting high-cost areas of the distribution system can promote cost-effective deployment of DR.[5]

C. Accommodating Distributed Resources

The policies that will best support the long-term deployment of distributed resources are the ones that enable the resources to be put to their most highly valued uses. In the main, this means that approaches that expose the value of the resources, and reward the resource owners for providing that value, should be implemented. In many cases, such policies rely on market mechanisms rather than on engineering prescriptions.

There are three major components of the US electric industry today in which distributed resources can figure prominently. How the "rules of the game" are set in each of these areas will affect the degree to which distributed resources are valued and deployed.

(1) *Wholesale Markets:* Wholesale markets should be designed to invite distributed resources, including demand-side price responses, to bid against supply on the trading floors of new

3. Arthur D. Little White Paper, *Distributed Generation: Understanding the Economics* (Cambridge, MA 2000), pp. 8-20.

4. Refer to RAP's report, *Distribution System Cost Methodologies for Distributed Generation*, companion to this paper.

5. See Section III.C.2., below.

electricity markets, and should permit DR to compete with transmission and generation investments to meet system needs;

(2) *Rates and Rules for Wires Companies:* Regulators should seek to send accurate price signals to customers and load-serving entities, and remove barriers to DR and demand-side resources; through reliability rules, rate designs for wires companies, and retail default service standards; and

(3) *System Benefits Programs:* Legislatures and regulators should create funding mechanisms for efficiency and load management investments, recognizing their reliability benefits, as well as the significant market barriers that still block their efficient deployment.

In this paper, we focus primarily on the first of these, but it should be recognized that there are linkages among them all. Sensible rate design and related practices and investments in long-run end-use efficiency will all have beneficial effects on wholesale markets.

1. DR in Regional Power Pools and New Electricity Markets

Wholesale markets must be structured to accommodate the potentially wide variety of distributed resources, both supply and demand, that are available. From the point of view of a system operator, distributed generation (DG) and load reduction look very much alike. The trick then is to devise mechanisms that enable both suppliers and end-users to discover and benefit from the value that their facilities can produce. As a general rule every effort should be made to expose the value of DR in the wholesale and retail markets to as many participants as possible.

The value of electricity varies from hour to hour, day to day, week to week, year to year. The variations in value flow from changes in production and delivery costs across time, themselves driven in large measure by changes in consumer demand across time. The intricate interplay between the supply of and demand for electricity has never been well exploited, in part because consumers and other market participants have long been insulated from cost changes by regulation and average-cost pricing. But substantial benefits to society and the environment can be captured by exposing those real, sometimes very high, costs of production, thus giving participants the opportunity to respond to them: by developing new technologies, curtailing load, and investing in more efficient end-uses.

Regulators need to structure markets and market rules so customers, retail sellers, distribution utilities, and other participants have an opportunity to realize the value of the services they can offer. Four policy reforms for accomplishing this should be implemented.

a. Demand-Side Bidding: Revealing the Demand Curve

Under traditional franchise regulation, the financial relationship between electric demand and 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 an expected load curve of customers who received only broadly averaged price signals. The "demand curve" in such circumstances was more an engineering concept than the revealed "willingness-to-pay" of the utility's customers.

Whatever the merits of this muted demand response in the franchise system, it has serious detrimental effects in competitive electricity markets that are established to efficiently balance demand and supply. As the recent price spikes, high prices, and reliability challenges of those new markets reveal, a principal lesson is that efficient energy markets simply cannot be built on such a foundation. There is ample empirical evidence that demand for electricity is not inelastic and that, at the high prices experienced in tight market situations, customers who have choices will respond by reducing demand or by 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.[6]

Whether 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 designed to reveal the customers' demand curve. The first step in this process is to require customers or their load-serving entities (LSEs) to place binding bids into the market under the same general conditions as generators placing supply-side bids. Bidding rules should permit, even encourage, load to bid at multiple price points, stating how consumption will vary as market prices change. In this way, market clearing prices and quantities will be determined as functions of the intersection of supply and demand (expressed, not expected) in specified periods.[7]

Offering one's on-site generation into the market is one aspect of demand-side bidding, insofar as it reduces the load to be served by central station dispatch. Consequently, a revealed demand curve informs the owners (or controllers) of distributed resources of the periods of high value. Armed with such knowledge, DR can be more efficiently deployed and operated.

---

6. See Cowart, Richard, *Efficient Reliability: The Critical Role of Demand-Side Resources in Competitive Electricity Markets*, The Regulatory Assistance Project, Spring 2001. Chapter V describes these issues in greater detail.

7. The general principle of demand-side bidding is straightforward, and all functioning markets have it in one form or another. In the electric industry, however, putting it into practice raises some complex challenges: for example, reforming the practice of using load-profiles (rather than interval metering) for the purpose of allocating peak and energy responsibilities among load-serving enmities. See *id*. for more detail on these and related matters.

b. Multi-Settlement Markets

A second needed reform in wholesale markets builds upon demand-side bidding, and extends the potential of demand and supply 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, but 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" bidding systems.

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.
(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 met by purchases from the spot market and are charged to suppliers and customers at spot market rates.

A "Two-Settlement" system similar to the one outlined here is now in operation in the PJM and New York regions; a similar system, termed "Multi-Settlements," is under development by the New England Independent System Operator for implementation in that region.

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.[8] Day-ahead

8. Peter Crampton and Robert Wilson (Market Design, Inc.), A Review of ISO New England's Proposed Market Rules, Executive Summary at 2.

prices are forecast, but with fairly low confidence.[9] 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 distributed 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.

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/or generate on-site, 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 for some reason the next day. Anyone who purchased supply in the day-ahead market now has a clear opportunity to profit by reducing consumption and selling back his power purchases into the spot market.

Importantly, demand-side sell-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 sell-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; this gives LSEs an incentive to predict their loads accurately, and to control their loads in times of thin margins and high prices.

This system of bidding will have at least two positive impacts:

- Improved system reliability by creating opportunities for demand-side as well as supply-side managers to meet the needs of the electric system during times of high prices; and

9. 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).

- Reduced market power by reducing supplier incentives to manipulate markets through the physical or economic withholding of assets (*e.g.*, by declaring units unavailable in the short term market).

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. They will also improve the strength and timing of price signals sent to 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 for mitigating the reliability challenges, prices spikes, and market power problems seen in wholesale power markets in the U.S. in recent years.

c. Opening Ancillary Services Markets to Distributed Resources

Electricity is a unique service in that production and consumption must be matched, for all intents and purposes, 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, controllable load and distributed generation can provide most balancing services just as well as controllable generation. And, as wholesale markets evolve to provide competition among generators, new opportunities can emerge for distributed resources to participate actively in providing reliability services to the power markets.

While FERC has stated 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.

Distributed resources, both supply and demand, offer obvious reliability benefits, but careful thought is needed to create market rules that will permit these 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 so that separate commercial entities 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.” 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 (RTOs).

As described earlier, there are eight ancillary (reliability) services that the owners of DR might want to sell. Providing ancillary services from distributed resources should involve a careful integration of generation and load response. Since fast services generally command higher prices than slower services, it is desirable to sell the fastest service possible. At times it may be faster to temporarily curtail load than to start generation. Load can be restored to service as additional generation is brought on line. It is also generally easier to incorporate energy storage in the form of thermal storage on the load side than it is on the power-supply side. Ten minutes of storage can be very valuable, as seen from the high prices paid for spinning reserves in Figure 1.

If ancillary markets are established so that distributed resources can participate actively (either directly or through aggregation), DR will benefit because it will receive revenue from the sale of ancillary services as well as from energy production. 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 generators. This will expand supplies and reduce horizontal market-power problems. Because ancillary services consume generating capacity, demand-side participation also improves overall resource utilization. When loads provide these reserves, generating capacity is freed up to generate electricity.

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, though that is the average response.

In contrast, 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 a the full 100 MW response each time.

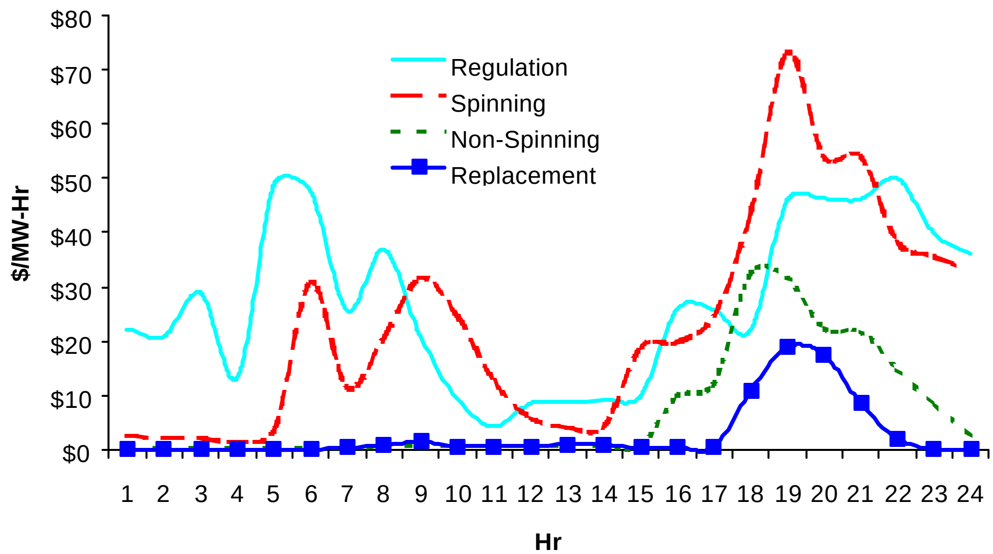

Figure 1 Real power reserve services requiring faster response command higher prices, on average, than do slower services in the California market as shown by this December 1998, weekday data.

d. Aggregation, Establishing Savings, and Other Market Rules

It goes without saying that the informational requirements of a complex, liquid, and geographically vast market in electricity are enormous – and absolutely critical to the successful functioning of that market. As the sizes of resources decrease, the relative costs of some information increase and could render uneconomic what would otherwise be a cost-effective facility or load-management program. As mentioned earlier, the aggregation of many small distributed resources under the control of a single manager – an LSE, distribution utility, or the like – offers a way of significantly reducing information and, in certain instances, operational costs while simultaneously providing the system operator with a large, highly reliable resource with which to balance supply and demand.

To the overall system, an aggregation of distributed supply and demand resources is indistinguishable from a larger central generating station, except insofar as the risk that that aggregated resource will be wholly unavailable for dispatch is much smaller than that of the central station. The difficulty, however, lies in establishing that the aggregator's resources do in fact produce power as bid or contracted for. The solution to this is straightforward enough, and does not require that performance be established in real-time, though of course performance must occur in real time.

First, as previously discussed, the multi-settlements system of supply and demand bidding and dispatch assures that no "phantom" load is included in the market-clearing process, and thus the potential problem of having to purchase load reductions for non-existent load is eliminated. LSEs (or aggregators) must pay for the energy provided under terms (prices, times, amounts) that satisfy their load bids in the day-ahead market.[10] LSEs now have a powerful incentive to curtail load or dispatch distributed generation when the real-time market price exceeds the price they are paying (whether under bilateral contract or as cleared the day before), so that any generation they free up can be sold back into the market at the higher prices. All that matters to the LSE is that the cost of curtailment or of DG dispatch is less than the real-time market price.

Of course, the load reductions or additional generation provided by the aggregator must in fact occur. While it might be nice to know at the time that the savings or generation occurred that they did in fact occur, it is not necessary; and the costs of such real-time metering and telemetry may very well be prohibitive.[11] But, because the consequences of non-performance can be significant – increased market prices and degraded reliability at certain times – it is important that the ISO have a high degree of confidence that the savings or generation are available when the aggregator asserts that they are. After-the-fact determinations of production (using on-site metering, for example) are therefore necessary, with the assessment of penalties if non-performance is established. The penalties, of course, should be stiff enough to discourage the withholding of resources, once offered, from the market. The same approach can be taken with respect to the provision of ancillary services.

What's important for regulators then is that policies and procedures promoting the market participation of aggregated loads be developed and implemented. Key among them include:

- *Simplified rules* for establishing production/savings (supply and load management).
- *Standardized metering techniques* that allow for reliable *post hoc* assessments. Interval metering, which links quantified amounts of production or savings to specified days and times, will be in many cases the preferred approach but, where the costs of metering outweigh the potential benefits, it should be possible to apply statistical methods to estimating the savings. Random testing will establish whether the savings are in fact generated.[12]
- *ISO communication protocols* that facilitate DR dispatch, to the extent that they are needed. Such protocols would allow for ISO interface directly with DR owners and with DR

10. The load and dispatch of bilateral contracts between suppliers and LSEs can be handled in several different ways by the ISO or market manager. Both their load and supply can be ignored for purposes of clearing the spot market, and they can be included and their bilateral contract regard simply as a "contract for differences." Either way, the bilateral arrangement should have no effect on the clearing of the spot market.

11. It's true, however, that the internet and other high-speed telecommunications systems have begun to bring these costs down significantly.

12. A difficulty with relying on assumptions about DR savings or production, rather than on metering, is that the system operator does not know for sure whether they in fact occurred. Payments might be made for "phantom" production. Pilot projects to test alternative approaches for measuring savings could be undertaken.

aggregators. Here it will be important for policymakers and system operators to bear in mind that DR's physical and operational characteristics (*e.g.*, ramp-up times, minimum run times, etc.) counsel against the imposition on DR of the same market participation requirements that are imposed on larger conventional units.

- *Recognition of the benefits of non-dispatchable DR.* Wind and photovoltaic resources are by their very nature both distributed and non-dispatchable. They operate only when the wind blows and the sun shines – times that very often coincide with times of high demand on the system, and thus they can have correspondingly high value. That value (apart from any environmental benefits) often exceeds the short-term market price for power, and it therefore behooves the system to promote, through longer-term contracting and other means, the deployment of such resources.[13]

Interconnection is not a subject of this report, but even so it shouldn't go altogether unnoted. The rules governing how distributed generation resources interconnect with the grid will determine whether and to what degree such facilities participate in the market. Policymakers must be sure that these rules assure safe and reliable interconnection, but are not so onerous as to inhibit cost-effective installations and dispatch.[14]

2. Other Areas for Corrective Actions

a.. Rates and Rules for Wires Companies

i. Transmission-Level Congestion Pricing

With the restructuring of wholesale markets has emerged the problem of network congestion and how to manage it. The traditional vertically integrated utilities accounted for transmission constraints when they made their daily operating (unit-commitment) plans. 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.

Constraints associated with transmission resources in the wholesaling (and consequently retail pricing) of electricity services were not been fully recognized in the past. Nor are they now. Transmission constraints impose significant costs on the system that are typically muted by a system uplift charge on all buyers. The variability of wholesale costs caused by such constraints needs to first be recognized in

13. Annual total production from wind and PV resources can be forecast with a high degree of accuracy. The same is true for next-day production, given expected weather conditions. Weekly and monthly projections, in terms of both total production and times of production, tend to be less reliable. Aggregating these resources for the purposes of system planning and dispatch greatly improve their reliability and value therefore.

14. An excellent examination of the barriers to interconnection faced by DR is *Making Connections*, by Brent Alderfer, Thomas Starrs, and Monika Eldredge (NREL/SR-200-28053, Boulder, CO, May 2000).

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, DR operators, and final consumers to manage loads during periods when transmission lines constrain access to the broader market.

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 is generally *not* related to the actual flows on lines. Congestion occurs when security-constrained dispatch requires modification of the economic dispatch. 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. It is this non-economic dispatch that results in locational price differences.[15]

In the long term, construction of new generators and transmission lines can reduce congestion. In the short term, system operators can treat congestion in two ways. They can mandate engineering solutions or they can use prices to let suppliers and consumers (*i.e.*, market participants) decide 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 transmission capacity were infinite). 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 curtailed depending on their firmness, with non-firm transactions cut before firm transactions 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 generators for any opportunity or out-of-pocket costs associated with this uneconomic dispatch. The costs thus incurred are then allocated to all transmission customers through an uplift charge. Although simple to implement, this approach is also economically inefficient because it fails to inform transmission users on the true

---

15. Losses also cause locational price differences but have a much smaller impact and are generally easier to deal with than congestion.

costs associated with their transactions. The absence of location-specific prices also robs investors of important information on where to locate new resources and what transmission projects to build.

Locational prices inform transmission users of the actual costs of transmission service and thereby promote economically efficient outcomes. Locational price differences reveal that the benefits of various energy supply and demand options depend not only on their temporal flexibility but also on their location. Distributed resources may have great value when they reduce load in the particular locations and at the particular times that congestion problems would otherwise arise. It is not necessary here to describe the various methods for calculating locational prices, which are in their details quite complex. Suffice it to say that distributed resources are admirably suited to exploit the differences in transmission costs on a system, and therefore it is imperative that RTOs and ISOs develop transmission pricing schemes that will give DR operators financial incentives to deploy their facilities or reduce consumption in the most cost-effective ways.[16]

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 will have the opportunity to compete to reduce the congestion and improve reliability. 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 "avoided cost" benchmark, replacing a system in which congestion costs are not revealed to customers, DR investments can compete on a fair basis with transmission and central generation options to provide reliability services in the load center.

ii. Retail Rate Design and Distribution Company Rate-Making

Although this paper is looking at ways to structure wholesale markets to accommodate distributed resources, it's worth adding a few words about actions that can be taken at retail to support DR, if for no other reason than that improving the economic efficiency of wholesale markets (which will have beneficial impacts upon DR) will be hindered by retail rate structures that discourage end-users from making efficient consumption decisions. Distributed resources offer cost-effective alternatives to purchasing energy from the grid, but they will not be exploited properly if their value is hidden from end-users.

Ultimately, retail rates provide the revenues needed to build and maintain the system. A wide variety of rate designs have been employed by utilities for decades, to serve a variety of purposes; but at least

16. By way of example, between April 1998 and September 1999, the average hourly price in the PJM area was $27.4/MWh. During this 18-month period, prices differed from location to location for 15% of the hours. During these congested hours, the maximum locational price difference averaged $19/MWh.

since the publication of James Bonbright's *Principles of Public Utility Rates* in 1961, the notion that rates ought to, among other things, be structured so as to promote economically efficient use of the electric system has been increasingly given the due that it deserves.[17] During the last two decades in particular, a large number of cost-reflective rate design reforms were adopted across the nation, including two-part (demand and energy) rates, seasonal rates, time-of-use rates, rate discounts for controlled load, and interruptible rates.[18] All of these initiatives were aimed at sending better price signals to customers: (a) to more accurately reflect the marginal costs of production and consumption; (b) to allocate more fairly the costs of the system; and (c) to improve reliability and lower overall system costs by removing inefficient subsidies and inspiring changes in demand patterns.

The nation's present focus on structuring electricity markets has drawn attention away from the underlying fact that rate design is still a critical function of regulation – almost all electricity is delivered on monopoly wires systems under regulated delivery charges; and the vast majority of energy sales are still made at regulated rates by regulated franchises or default service providers.[19] For these consumers, rate designs that reflect the economic costs of production and delivery should not be abandoned. There are several other options that could significantly improve customer price-responsiveness; utilities and state regulators in any jurisdiction facing reliability concerns should examine the following:[20]

- *Time-differentiated default rates and transitional price caps:* Customers may prefer rate stability to free-wheeling volatility, but they do not require a single rate for every hour of the year.[21] Considering the enormous costs and reliability concerns associated with seasonal peaks, any annual price caps adopted as part of a restructuring plan or utility rate freeze should include meaningful differences between peak and off-peak consumption. On an average annual basis,

17. The call for the application of economic principles to the challenge of rate design had, of course, been heard much earlier. See, for example, Boiteaux, Marcel, "La Tarification des demandes en pointe: application de la théorie de la vente au coût marginal," 58 *Revue générale de l'électricité*, 1949 (updated as "Peak-Load Pricing," translated by H.W. Izzard, 30 *Journal of Business of the University of Chicago*, 1960, p. 157).

18. An EPRI survey in 1990-91 reviewed more than 1,000 "innovative rates" of these types, offered by 135 major utilities. EPRI, *Survey of Innovative Rates, 1991* (Palo Alto, April 1992) (In three volumes).

19. As of June, 2000 about 98% of the national load is still provided by incumbent franchises. XENERGY estimate, as reported in Restructuring Today, Monday, June 26, 3000 at 3.

20. This is by no means an exhaustive list. Many other innovations in rate design ought to be considered in the proceedings recommended above. Some of these innovations are relatively new, such as net metering; others have long been used, but may need revival and adjustment in light of current technologies and changes in market structure. For example, regulators in Vermont are moving away from seasonal rates, which have had a very large and beneficial effect on capacity factor and peak load exposure for the state's utilities over two decades. The move is based upon changes in the regional wholesale market, where winter power costs have moderated. But seasonal rates have not been replaced by any new rate design to reflect the newly-volatile, high costs of consumption in the wholesale market at peak periods, both summer and winter.

21. Seasonal rates and time-of-use rates have long been a feature in many jurisdictions, but their usefulness in controlling peak load has been eroded by inattention and the assumption that market prices would soon take over.

"the default price" might well be the same, but reliability will be improved when consumers see the cost of maintaining peak in the rates they pay during peak periods.

- *De-averaged buy-back rates:* Despite the appearance that distribution costs do not vary directly with usage, in fact they do – particularly when viewed in the longer run or when demand presses up against the limits of capacity. At times of capacity constraint, the marginal costs of delivery rise very steeply: they are, in fact, the costs of new investment in wires, transformers, and substations. Moreover, as with transmission, the marginal costs of distribution can vary significantly by time and location.

  Where the marginal costs of distribution are high, the utility has a strong incentive to invest in less costly means of providing service: end-use efficiency, distributed generation, and load management, for instance. This is particularly true where, as in most areas, the retail rates for distribution service are averaged, and marginal on-peak costs exceed marginal revenues. In such circumstances, utilities have a very palpable profit motive to reduce costs. Customers, in contrast, do not. They are not being given price signals that reflect the full marginal costs of service, at least at times of peak, and consequently their incentives to invest in distributed resources are muted. And, if they are paying fixed fees for distribution, the incentives are non-existent altogether.

  One response is to de-average distribution prices, according to location. However, assuming that the geographic de-averaging of prices is not possible, alternative approaches for promoting economically efficient outcomes must be developed.[22] One such approach is the geographically de-averaged "buy-back" credit. The utility would establish financial credits for distributed resources installed in a given area. The credit amount would be a function of the distribution cost savings generated by the distributed resources. Credits would be limited in duration and magnitude, in order to match the timing and need for distribution system reinforcements. For example, credits might be available to the first 20 MW of distributed resources installed in the next year because, after that period, loads are expected to have grown to the point that distribution line upgrades are unavoidable. The dollar amount of the credits should, at most, equal the value (savings) derived from deferring the distribution upgrade. Credits would also vary by location of the distributed resources. Credits would be highest in areas of greatest need and would be as low as zero in low-cost areas.[23] For example, customers in an area with 20¢

22. To the economist, differentiating prices according to geographic cost characteristics is no different than doing so according to time of use. However, in light of the potentially very great differences in rates from area to area, the administrative complexity of the rate structure, and universal service considerations, we are unlikely to see geographically de-averaged rates any time soon.

23. Variations of the de-averaged distribution credits could be a sliding scale standby rate or a hook-up "feebate." For example, stand-by rates could be on a sliding scale ranging from high to negative. Negative stand-by rates, which look like distribution credits to customers, would be charged in high-cost areas. A hook-up feebate would be a

(continued...)

distribution costs might be offered a 15¢ credit.[24] This would certainly produce a strong economic incentive for customers and others to invest in distributed resources. Because the credit is 15¢ instead of the 20¢ the utility would incur to upgrade facilities, there is an opportunity for savings to be shared.[25]

- *Revenue caps, not price caps or fixed charges, for wires companies:* At first blush, the rate design for wires company services would not appear to raise significant reliability concerns. But it does. In a price caps environment, the "lost profits" problem will continue to undermine broad-based energy efficiency improvements in end-uses throughout the grid. This will raise overall consumption levels, erode reserve margins, and put increased stress on distribution and transmission systems. A system of high, fixed charges, as proposed by many utilities today, appears to address this problem, but it causes problems of its own. In this case, the wires company no longer has an incentive to promote high throughput. But, having paid a high fixed charge for access, the customer faces a much lower rate for incremental consumption. Since incremental consumption drives peak, high peak load prices, and reliability problems, high fixed customer charges can promote inefficient consumption and degrade reliability.

  The solution to both of these problems is a performance-based, per-customer revenue cap.[26] It rewards a firm for increases in operating efficiency, while making it indifferent to the volume of throughput over its wires. Since, in the short run, a distribution company's costs vary more closely with the number of customers it serves, than with throughput, a per-customer revenue cap would produce annual revenues that more closely track annual costs. To the utility, a per-customer revenue cap looks just like a fixed-price rate structure, and it removes the company's disincentive to support customer installations of efficiency and other distributed resources. However, the revenue cap enables prices for end-users to be set on a usage basis, enabling them to make consumption decisions and alternative energy investments that are, in the longer term, more efficient. In addition, if the per-customer revenue cap is modified by performance objectives, the resulting PBR plan can adjust rate levels automatically to encourage the utility to pursue cost-effective distributed resource options, and lower the overall cost of the distribution system. The resulting savings can be shared between the company and its customers.[27]

23. (...continued)
revenue-neutral charge that collects from customers installing distributed resources in low-cost zones and pays to customers who install distributed resources in high-cost zones.

24. Demand-side resources are so much less costly that the winning bid prices would likely be far below 15¢.

25. Moskovitz, p. 24.

26. See David Moskovitz, *Profits and Progress Through Distributed Resources*, NARUC, 2000 pp.16-18, 20-22, and Frederick Weston, *Charging for Distribution Utility Services: Issues in Rate Design*, NARUC, 2000.

27. A rate plan with most of the attributes described here is now in effect for Pacificorp in Oregon, and (before the reliability and financial crises of 2001) was being considered by the major distribution companies in California.

b. Promoting End-Use Efficiency

Decision-makers addressing the reliability problems of emerging wholesale power markets may find that, because they are focusing on the problems of peak load, they are drawn particularly to load management solutions. Demand-side bidding, price-responsive load, and "dispatchable load" ancillary services are very important resources to electricity systems and to reliability managers. Broad-based energy efficiency options may thus be overlooked, despite their economic and reliability benefits. But this would be a grave mistake, since the market barriers to investment in long-term end-use efficiency that led to utility-sponsored demand-side management (DSM) programs in the 1980s and '90s still exist. There's nothing about industry restructuring to have changed that.

The present challenge is to create mechanisms for delivering broad-based efficiency measures to electric networks. The first and most obvious opportunity would be to reinvigorate the practice of utility Integrated Resource Planning, particularly in those franchises that are not likely to face retail competition in the near-term – which, for the next decade, may well be half of the nation. While logical, this may not be a promising avenue. Utilities in regions not open to competition are nevertheless anticipating the *possibility* of retail competition in the future, and will likely remain reluctant to invest heavily in efficiency measures.[28] Fortunately, regulators and legislators have other models to consider. Among the most promising are:

- *System Benefit Funds:* Broad-based wires charges can support efficiency and load management measures that enhance system reliability and lower market prices
  In the absence of utility funding of efficiency programs in rates under integrated resource planning, it is possible to support them through broad-based wires charges, assigned to the electric bills across broad classes of customers. As a non-bypassable charge, no competitive provider is disadvantaged by the collection mechanism. At least 11 states, including California, Wisconsin, Ohio, New York, and Illinois have established statewide funding mechanisms for efficiency programs, supervised by state agencies with a mandate to improve reliability and save energy cost-effectively.

- *The Energy Efficiency Utility:* One important variant on the statewide public benefits fund is the "energy efficiency utility," which is awarded a franchise in order to deliver efficiency services to customers across a state or region. The first such utility was chartered by the Vermont Public Service Board, with a statewide franchise, supported by a wires charge in each franchise territory in which it delivers services. It was designed to eliminate the conflict of interest that wires companies have with respect to most efficiency services, and is supported with funds that formerly went to power company DSM programs.

28. On the other hand, a utility that invests wisely in efficiency in its home territory may be able to reap the benefits of sales at high prices in wholesale markets and save on purchase power costs.

- *System Benefit or Uplift Charges at the Power Pool Level:* The wholesale markets could be designed to capture large consumer savings through broad-based market transformation or energy efficiency programs without much difficulty. One option is a wires charge devoted to investment in energy efficiency that pays for itself. For example, an investment in energy efficiency funded through a wires, or uplift, charge equivalent to 5 mils per kWh will reduce average wholesale prices by 10 mils. This is a direct net benefit to all electricity users in the form of lower wholesale prices, not to mention reduced air pollution. 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.

IV. CONCLUSION

If we hope to exploit the wider range of reliability and energy services that distributed resources can provide, important policy changes need to be made. This paper describes the key set of them. They are not technology-prescriptive – they do not ordain winners and losers – but instead they aim to give incentives to market participants to develop innovative, more reliable, and less risky methods of meeting the nation's demand for electricity. These policies include the following:

- Demand-side bidding and multi-settlements;
- Demand response (participation of load management in spot markets);
- Opening the ancillary services market to DR;
- Resource aggregation and management;
- Increasing market liquidity;
- More economically efficient transmission and distribution rate design; and
- Public benefits programs, including funding mechanisms, in support of investment in long-term end-use energy efficiency.

Viewed another way, these policies identify market, rather than engineering, mechanisms that can expose the value of distributed resources. Once uncovered, we leave it to the many thousands of dedicated and creative people around the world to find the best ways of capturing that value.