Response Program Evaluation submitted to FERC, December. Available at www.ISO-NE.com.

Ruff, Larry E., 2002, "Economic Principles of Demand Response in Electricity" report to the Edison Electric Institute, October.

Smith, V. and L. Kiesling, 2005, "A Market-Based Model for ISO-Sponsored Demand Response Programs: White paper prepared for a Multi-Client Study: A Critical Examination of Demand Response Programs at the ISO Level: End Goal, Implementation, and Equity" Center for the Advancement of Energy Markets, Distributed Energy Financial Group, August.

Spulber, D.F., 1992, "Capacity-Contingent Nonlinear Pricing by Regulated Firms" *Journal of Regulatory Economics* 4(4):299-320.

Stoft, Steven 2004, *Prepared Direct Testimony before the Federal Regulatory Energy Commission* FERC Docket No. ER03-563-030, August 31.

Summit Blue Consulting, 2004, "Evaluation of the Energy-Smart Pricing Plan: Final Report" prepared for Community Energy Cooperative, February.

Summit Blue Consulting, 2005, "Evaluation of the 2004 Energy-Smart Pricing Plan: Final Report" prepared for Community Energy Cooperative, March.

Taylor, T., P. Schwarz, and J. Cochell, 2005, "24/7 Hourly Response to Electricity Real-Time Pricing with up to Eight Summers of Experience" *Journal of Regulatory Economics* 27(3):235-62.

U.S. Department of Energy (DOE), 2003, "Report to Congress: Impacts of the Federal Energy Regulatory Commission's Proposal for Standard Market Design" DOE/S-0130, April 30.

Violette, Daniel and Michael Ozog, 2003, "Mass-Market Demand Management Offerings: Evaluation Methods Assessment and Results" presented at the International Energy Program Evaluation Conference, Seattle WA, August.

Violette, Daniel, Rachel Freeman and Chris Neil, 2006, "DRR Valuation and Market Analyses, Volume II: Assessing the DRR Benefits and Costs" task status report to the International Energy Agency Demand-Side Programme: Task XIII: Demand-Side Resources Project, January 6.

Woo, C.K., 1990, "Efficient Electricity Pricing with Self-Rationing" *Journal of Regulatory Economics* 2(1):69-81.

York, Dan and Marty Kushler, 2005, "ACEEE's Third National Scorecard on Utility and Public Benefits Energy Efficiency Programs: A National Review and Update of State-Level Activity" American Council for an Energy Efficient Economy: ACEEE-U054, October.

APPENDIX A. ORGANIZATIONS THAT PROVIDED INPUT ON RECOMMENDATIONS

American Council for an Energy-Efficient Economy
American Public Power Association
Apogee Interactive, Inc.
Arkansas Public Service Commission
Battelle-Pacific Northwest National Laboratory
BP Solar
California Department of Water Resources State Water Project
California Energy Commission
California Public Utilities Commission
Constellation Energy
Consumer Energy Council of America
Cornell University
Demand Response and Advanced Metering Coalition
Distributed Energy Financial Group
Duke Power
East Kentucky Power Cooperative
Edison Electric Institute
Energy Connect Inc.
Grid Services, Inc.
Hunt Technologies, Inc.
Idaho Public Utilities Commission
Invensys Controls
ISO New England, Inc.
Itron
Louisville Gas and Electric
M.Cubed
National Rural Electric Cooperative Association
New York State Department of Public Service
PJM Interconnection, LLC
San Francisco Community Power Cooperative
Solar Turbines, Inc.
Southern California Edison Company
Steel Manufacturers Association
SUEZ Energy NA
The Cool Solutions Company
The Stella Group, Ltd.
U.S. Department of Energy—Building Technologies Program
United States Demand Response Coordinating Committee
Utilipoint International, Inc.
Utility Economic Engineers

APPENDIX B. ECONOMIC AND RELIABILITY BENEFITS OF DEMAND RESPONSE

This Appendix provides a more detailed conceptual discussion of the economic and reliability benefits of demand response than was included in Section 3. First, short-term market impacts are described, drawing on economic theory to show how demand response can result in improved economic efficiency, and distinguishing how these benefits are manifested under different market structures. Next, long-term economic benefits from avoided capacity investments are discussed along with issues in designing and implementing programs designed with this goal in mind. Differences in how short-term and long-term economic benefits are realized and passed on to consumers are then described for vertically integrated utilities and regions with ISO/RTO spot markets. Finally, reliability benefits are described along with concepts used to value them.

Short-Term Market Impacts: Supply Costs and Market Prices

This section provides a detailed discussion of how customer load reductions lower energy supply costs in the short term. First, the basic source of short-term market benefits—improved economic efficiency brought about by allowing consumers to make electricity usage decisions based on marginal, rather than average, supply costs—is described. Differences in how these benefits are manifested in regions with differing market structures are then discussed.

Societal Benefits

In evaluating policies or structural changes that impact how markets work, economists distinguish between societal gains, which benefit everyone, and financial flows that involve gains by some at the expense of others, called transfers. In the absence of a way to weigh the relative impact on individuals of gains and losses (i.e., a change in utility), economists argue that policies should primarily be judged on their net outcome, which is defined by the level of societal benefits (see the textbox below).

Demand response produces societal benefits, which are resource savings, by reducing the gap between time-varying marginal supply costs and retail electricity rates based on average costs. Economic theory asserts that the most efficient use of resources occurs when consumption decisions are based on prices that reflect the marginal cost of supply. In a competitive market, this is defined by the intersection of a good's supply and demand curves (see Figure B-1). In electricity markets, the marginal electricity supply curve is constructed by ordering generators from lowest to highest operating costs (often referred to as "merit order").[68] Due to the technical characteristics of electricity generation equipment, the supply curve—the upward curving line in Figure B-1—tends

[68] Certain generators may be required to run, regardless of their marginal operating costs, to maintain reliability in areas with constrained generating and/or transmission capacity, which limits the ability of least-cost resources to serve local demand.

to increase very steeply at its upper end.[69] This means that when demand approaches the industry's installed capacity, each additional increment of demand imposes increasingly more cost than the previous one. In other words, the marginal cost of electricity becomes most sensitive to changes in demand when demand is already high.[70]

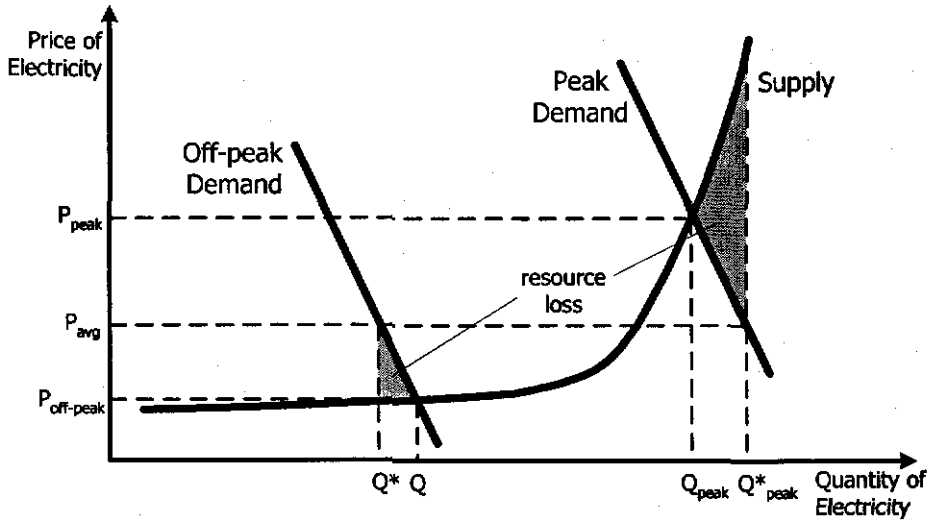

Figure B-1. Inefficiencies of Average-Cost Pricing

Like most goods, the demand for electricity exhibits declining marginal value (i.e., the marginal value of additional consumption declines as consumption increases). Electricity demand is characterized by a downward-sloping line, regardless of how electricity is priced. But, if the price that consumers pay never varies, demand appears to be perfectly inelastic, and is characterized by a vertical line. Moreover, consumers' demand for electricity also depends on the time of day, with more usage typically occurring during the "peak" afternoon and early evening hours and less at other times. This phenomenon is driven by the economic activity of businesses and residential customer lifestyles and usage patterns, but is also influenced by electricity rates that are the same throughout the day. For simplicity, the two lines labeled "peak" and "off-peak" in Figure B-1 represent consumer demand.

The most efficient pricing and usage of electricity is determined by the intersection of the supply and demand curves in Figure B-1. In other words, during off-peak periods, the efficient price of electricity should equal $P_{off\text{-}peak}$ and consumers would use an amount of

[69] The long, flat portion of the electricity supply curve represents "base-load" power plants, such as nuclear, hydroelectricity and coal plants that have very low operating costs and are run most hours of the year. Base-load plants are typically large with similar characteristics. The steeply inclining portion of the supply curve represents "peaking" plants that are used to meet peak demand needs and may be run only a few hours per year. These plants are typically natural gas- or oil-fired combustion turbines that are less expensive to build than most base-load technologies but have higher operating costs. Peaking plants are typically smaller units with varied operating characteristics.

[70] High demands do not always lead to high prices. If the entire portfolio of capacity is available, then the marginal unit may be relatively low cost. The steepest part of the supply curve is encountered when demands are especially high (e.g. a heat wave) or generation is short due to forced outages, or both.

electricity equal to Q, and during peak hours, the efficient price should equal P_{peak} and consumers would use Q_{peak} units of electricity. However, most consumers currently pay electricity tariffs that reflect average, rather than marginal, electricity supply costs; this is represented by P_{avg} in Figure B-1. Actual usage therefore reflects the intersection of the demand curves with this average price, resulting in less than the social optimal usage in off-peak periods (Q*) and more than the social optimal usage in peak periods (Q^*_{peak}) relative to the optimally efficient system.

Distinguishing Societal Benefits from Rent Transfers

Economists make a distinction between *transfers*—the benefits of a policy initiative that amount to gains for some at the expense of others—and *social welfare gains* that inure to society as a whole. Social welfare gains are desirable because they derive from efficiency improvements that benefit all market participants. These benefits provide a strong rationale for policymakers to invest consumers' money in initiatives to realize such gains. Transfers result in some market participants being better off than others. In the case of demand response, lower market prices reduce revenue to suppliers and lower costs to consumers. The economists' task is to quantify the relative marginal gains and losses to the individuals involved.

Some economists caution that treating market price reductions as benefits is misleading, and may result in policies that undermine, rather than enhance, market efficiency (Ruff 2002). Specifically, they contend that using the bill savings from price reductions, which largely amount to transfers, to justify demand response incentive payments to customers actually raises electricity prices in the long term. They contend that merchant generators count on the profits (called scarcity rents) realized when prices are high to recoup their capital costs and achieve the rate of return their investors require. If these profits are reduced because policymakers use them to justify customer curtailment incentives, then investors will become more skeptical and require higher returns, which, the argument concludes, results in higher prices in the long run.

This is the basis for many of the objections to allowing customers to bid load curtailments as resources into ISO/RTO spot markets, called "demand bidding as a resource." However, other economists contend that if demand response moves the wholesale market to greater economic efficiency and the result is a more appropriate supply and demand balance, then the elimination of those artificial rents to generators corrects a market distortion and prevents investments that are not needed based on how customers value electricity.

Another objection to demand bidding raised by some economists is their claim that customers on default service have no right to the energy, since the utility rates require that it be served, but do not give the customer any contractual rights to that supply. This could be corrected by requiring that in order to bid curtailments into spot energy markets, the customer would have to demonstrate that it has contractual rights to that power. As an alternative, these critics propose "self-financing" demand response whereby the inherent savings from avoiding paying high market prices is the inducement for customers to curtail, and no payment has to be made to achieve that result (Braithwait 2003).

These arguments have only been raised for demand response programs that allow customers to offer curtailments as resources in centrally organized spot markets. Yet, substantially the same transactions characterize demand bidding and CPP programs run by vertically integrated utilities.

Economists refer to the inefficiencies that arise when retail prices do not reflect marginal supply costs as "dead-weight losses" or resource losses (i.e., the loss of societal welfare when resources are not used optimally). The resource losses from average cost pricing are illustrated by the shaded triangles in Figure B-1. In the off-peak period, electricity that

would have value to consumers if it were priced according to its marginal supply cost is not consumed—this represents a loss to society in economic activity that would have occurred but did not. In the peak period, consumers that do not pay the full marginal cost of power consume excessive amounts of electricity at a cost in excess of the value it provides them. Because this occurs at the steeply inclining portion of the electricity supply curve, these costs can be substantial.[71]

The short-term market-impacts benefit of demand response lies in reducing or eliminating this resource loss, thereby improving net social welfare. The combined resource loss from all peak and off-peak hours—and thus the potential for short-term demand response benefits—depends on how widely average and marginal electricity costs vary. For example, in a tightly constrained market, where peak demand is often very close to supply limits, the potential short-term efficiency benefit from implementing demand response can be substantial.

Supply Cost and Market Price Impacts in Regions with Differing Market Structures

Short-term market impacts are illustrated for vertically integrated utilities in Figure B-2. The supply curve typically reflects the utility's supply costs, including its own generation plants and any incremental wholesale power purchases. If demand is forecast to be Q, then a demand reduction that moves consumption to Q_{DR} results in an avoided utility supply cost equal to the shaded area in Figure B-2.

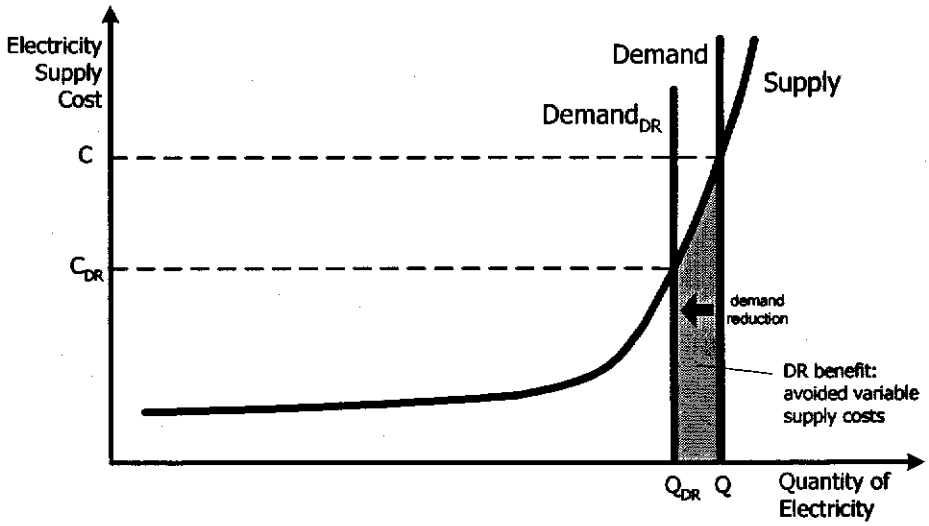

Figure B-2. Impact of Demand Response on Vertically Integrated Utility Supply Costs

The same load reduction produces more extensive impacts in regions with organized wholesale markets because of the way these wholesale markets are designed. The supply curve is developed by arranging generators' offer bids in merit order from lowest to

[71] Electricity pricing that does not reflect supply costs results in societal losses both when costs are high, and when they are low. However, the extent of these losses is greater at elevated supply costs, and therefore correcting prices in these periods has captured the attention of policymakers and market designers.

highest. Because of competition among generators, generators' offer bids reflect their marginal operating and maintenance costs and in some circumstances additional margins to recover fixed costs. LSEs also bid their expected load requirements into the market, producing a demand curve.[72] The bid price of last generator needed to serve the LSE's purchases sets the market clearing price for the whole market. This means that a demand reduction from Q to Q_{DR} not only provides the avoided variable cost savings observed for vertically integrated utilities (the shaded area to the right in Figure B-3), but it also lowers the price of all other energy purchased in the market. This second market impact, represented by the shaded rectangle in Figure B-3, is dependent on the level of price reduction—the difference between P and the new price P_{DR}—and the amount of energy bought in the applicable market. LSEs typically commit their expected energy requirements with a mix of bilateral forward contracts with generators and purchases in day-ahead and real-time markets. This is represented by the dotted line in Figure B-3. The extent of customer savings from price reductions thus depends on how much energy is purchased in spot markets.[73]

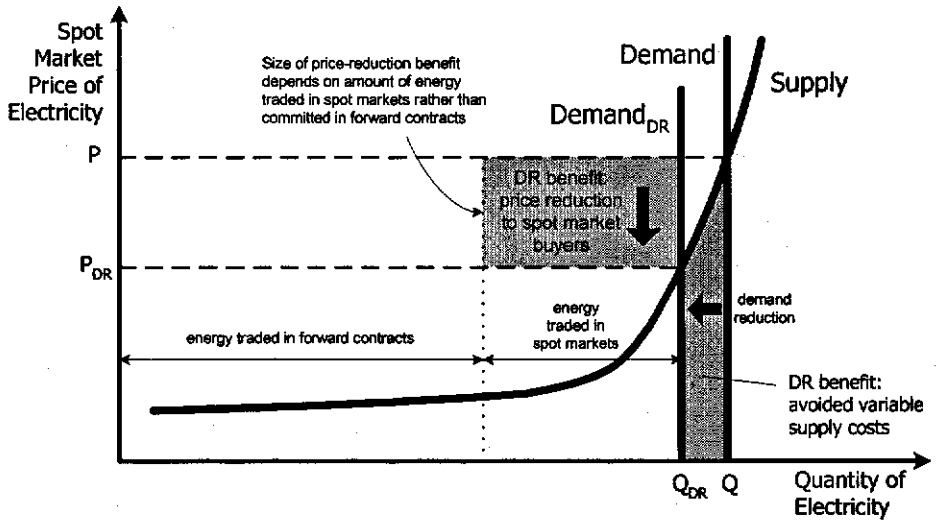

Figure B-3. Impact of Demand Response in Regions with Organized Wholesale Markets

In regions with organized wholesale markets, if, over time, customers routinely respond to high prices by curtailing or shifting loads, then additional, longer-term savings will result. Thus, if demand response consistently reduces market prices and volatility, bilateral contract prices will also drop over time, as reduced price risk in day-ahead and real-time markets pushes longer-term contract prices down. This is because LSEs may be willing to pay less for hedged forward contracts and will buy instead from the spot market if generators do not offer lower forward contract prices. In this way, lower energy

[72] In this example, demand is represented by a vertical line for simplicity (i.e., it is presumed to be fixed). Currently, most LSEs bid fixed quantities of electricity in spot markets, so this characterization is appropriate.

[73] In New York, a state with organized wholesale markets and retail competition, over 50% of electricity is traded in day-ahead and real-time spot markets, with the rest settled in forward contracts. In New England, about 40% of the electricity volume is traded in ISO-NE's spot markets, with about 60% committed in forward contracts.

prices resulting from short-term demand response market impacts can eventually extend to the entire market.[74]

Long-term Market Impacts: Capacity Benefits

The long-term market impacts of demand response hinge on reducing the *system peak demand*—the highest instantaneous usage by consumers in a particular market. Reducing system peak demand can avoid or defer the need to construct new generating, transmission and distribution capacity, resulting in savings to consumers. This applies for both vertically integrated utilities and organized wholesale markets, although capacity costs are allocated differently. This benefit can be specifically elicited from customers through capacity-based demand response programs (e.g., DLC, I/C rates or ISO/RTO capacity based programs) or may result from consistent load reductions from price-based demand response options (e.g., RTP). For example, in a capacity-based demand response program, load reductions timed to reduce load from a level that otherwise would have established the system maximum demand can yield large benefits for all consumers. Historical system maximum demand, adjusted for planned reserves, establishes ongoing generating capacity requirements, usually on an annual or semi-annual basis. For example, if the maximum demand served in a control area during the past summer was 5,000 MW, then that demand would serve as the basic capacity target for the next summer, to which an additional reserve margin (e.g., 18%) would be added.[75] If the existing infrastructure were insufficient to serve the resulting 5,900 MW capacity requirement, additional capacity would be necessary. Since generating capacity is expensive, ranging from about $50,000 to over $100,000 per MW-year (depending on the type and location of generating units), demand response that displaces the need for new infrastructure can produce substantial avoided cost savings.

Demand response programs designed to reduce capacity needs are valued according to the marginal cost of capacity. By convention, marginal capacity is assumed to be a "peaking unit", a generator specifically added to run in relatively few hours per year to meet peak system demand. Currently, peaking units are typically natural gas turbines with annualized capital costs on the order of $75/kilowatt-year (kW-year) (Orans et al. 2004, Stoft 2004). Thus, if demand response programs avoid 100 MW of generating capacity, the avoided capacity cost savings would be $7.5 million per year in this example. If the total program costs were $50/kW-year, including incentive payments to participating customers, then other customers realize the rest as savings (e.g., $2.5 million per year in this example), which may eventually be reflected in lower rates and bills. As long as there is some sharing of benefits, all customers benefit from others' participation in a capacity demand response program.

[74] Whether or not savings from short-term market price impacts and reduced forward contract prices brought about by incentive-based demand response programs should be treated as societal benefits is a subject of controversy (see the textbox on "Distinguishing Societal Benefits from Rent Transfers", earlier in this Appendix).
[75] Reserve margins vary in electricity markets across the U.S., but are typically 15-18%.

Transmission and distribution system capacity investments are also capital-intensive, and demand response that reduces local maximum demand in areas nearing infrastructure capacity can also provide significant avoided cost savings.

Realizing Capacity Benefits: Establishing and Reducing System Peak Demand

Capacity-based demand response programs are designed to replace generation investments and participants receive up-front capacity payments tied to this avoided cost. To realize this benefit and justify making the capacity payments, system operators must be able to dispatch curtailments that actually avoid building new capacity. This is accomplished in one of two ways: (1) predicting when system peak demand will exceed historic levels and dispatching load reductions accordingly or (2) dispatching curtailments when a designated peaking generation unit would otherwise be in service.

Dispatching demand response to avoid increasing system peak demand involves predicting when peak demand is likely to exceed historic levels absent any curtailments. Electric systems are generally either winter or summer peaking, meaning that annual demand is seasonal. However, demand can exceed historic peak levels several times during the peak season, which may span several months. To ensure that a capacity program truly does reduce peak demand, operators may need to dispatch the program several times during the peak season to account for forecast error. For participating customers, multiple curtailment obligations can be burdensome. To improve the attractiveness of capacity programs to customers, limits are sometimes are placed on how many curtailments can be called in a particular season.

The alternative method is to dispatch capacity-based demand response programs when an existing plant designated to meet peak demand would be needed to serve expected demand, absent any curtailments. This practice is somewhat more straightforward in regions with organized wholesale markets because transparent market rules direct dispatch operations. However, vertically integrated utilities have similar unit dispatch rules that could be used. Here too, limits may be placed on how frequently curtailments are called for.

Both methods of dispatching demand response to realize capacity value require provisions for periodic testing of customer response as well as penalties for non-performance. Testing is necessary to certify that customers truly have the capability to deliver the contracted curtailments on an on-going basis. Penalties serve to reinforce their obligation to be available and deliver load reductions when called. However, establishing appropriate penalty levels can be challenging. Increased penalty levels make demand response commitments more reliable and more valuable to the system operator, but are likely to reduce the amount of demand response committed by customers.[76] Program designers must balance the attractiveness of the program to customers against the potential consequences of forced outages that affect a large number of customers at costs well in excess of the avoided cost payment participating customers receive.

Because the avoided capacity cost savings calculation is prospective, so is the value of a capacity-based demand response program. This raises issues in forecasting the timing of system peak demand, or the highest 10-30 load hours of the year, so that calls for demand reductions actually moderate system maximum demand as designed. Since forecasting involves errors, program administrators/sponsors must make provisions to ensure the

[76] One useful strategy may be to recruit larger numbers of customer participants by dropping or reducing penalties for non-performance. Even though each customer is a less reliable source of demand response in the absence of penalties, the larger number of participants could increase the total expected demand response. The adoption of such a strategy would require evaluation of accumulated experience on the effect of various levels of penalties on customer performance.

demand response program is called often enough to effectively lower the forecast of system peak demand (see the textbox above).

Timing and Distribution of Market Impacts of Demand Response

Differences in market structure influence the timing and distribution of short-term and long-term market impacts of demand response in important ways. These differences are illustrated in this section by tracing the market impacts and resulting benefits of demand response in two types of market structure: 1) "vertically integrated systems", in which a vertically integrated utility with a retail monopoly franchise engages in some wholesale market transactions but operates in a region without an ISO or RTO, and 2) regions with organized wholesale markets in which ISOs/RTOs administer spot markets and retail competition is enabled at the state level. These illustrative combinations of retail and wholesale market structures reflect the current situation in many states or regions, although other retail/wholesale market structures are prevalent in the U.S.[77]

In this section, the examples suggest that the market impacts of demand response within organized spot markets produce benefits in a *shorter* timeframe than those for a vertically integrated, monopoly utility.

Market Impacts of Demand Response for Vertically Integrated Utilities

Vertically integrated utilities are responsible for making capacity investment decisions (whether to build new generation itself or to purchase supply contracts from other sources such as independent power producers), subject to regulatory oversight and approval, and for planning and operating the electricity grid and ensuring reliability. Retail rates are determined administratively, based on the average cost of supplying all three major facets of electricity production and delivery—production, transmission and distribution—and expected sales volumes. Embedded in retail rates are marginal costs to supply power, such as fuel, operating and maintenance costs, as well as a return on investment for un-depreciated utility-owned generation.

The economic impacts of demand response for a vertically integrated utility operating with a retail monopoly franchise are depicted in Figure B-4. Short-term demand response benefits may be traced as follows:

- Depending on the timing and type of demand response option, customers' load changes may be integrated into the utility's scheduling and dispatch decisions on a day-ahead or near-real-time basis.
- Changes in load (e.g., reductions in usage during high-priced peak periods) offset a portion of usage that otherwise would have been met by production from high-

[77] For example, utilities in some states are still vertically integrated and retain a retail monopoly franchise but are part of an organized regional wholesale market administered by an ISO or RTO (e.g., some parts of MISO, Vermont).

operating-cost power plants or purchases during the load response event (see Figure B-2).[78]

- This lowers the average variable electricity cost, which should be manifested eventually as customer bill savings through lower regulated electricity rates.

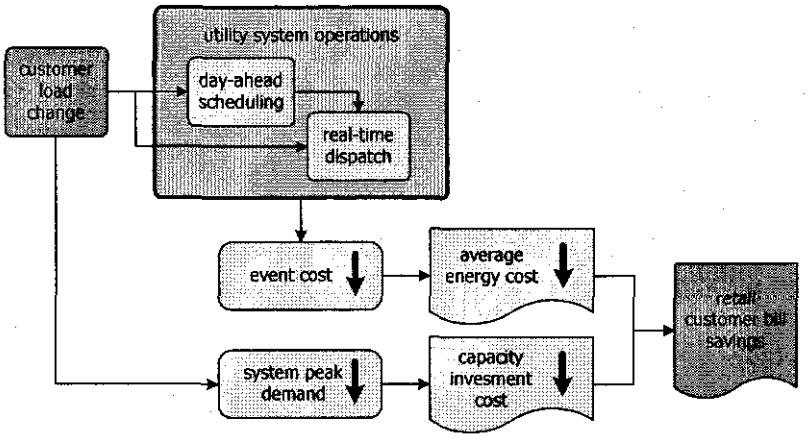

Figure B-4. Market Impacts of Demand Response for Vertically Integrated Utilities

The utility's return on capacity investments is recovered separately from its marginal costs to produce or purchase electricity and operate the electric grid. Thus, in vertically integrated systems, in the absence of a mechanism to reveal marginal capacity or reliability costs in unit operating costs, the short-term market impacts of demand response are limited to efficiency improvements in operating costs (including energy production and purchase costs) alone.[79]

In the long term, demand response that reduces peak demand growth directly averts the need for utilities to build more power plants, power lines and other capacity-driven infrastructure or to buy new capacity and energy from other suppliers (see Figure B-4). Because capacity investments are usually fully recovered—along with a pre-established return on investment—through higher retail electricity rates, these long-term benefits are realized over a multi-year period and can result in significant savings to consumers.

In vertically integrated, stand-alone utility systems, demand response is most useful to improve generation and transmission asset usage, avoid new capacity construction or purchases, and create more flexibility to assure reliable system operations. This influences the types of demand response programs preferred by vertically integrated utilities, as well as how they value and compensate demand response program participants.

[78] The converse is true for increases in load at times when the marginal cost of electricity is lower than the average retail price.

[79] Some utilities quantify the marginal value of reliability in their RTP tariffs quoting hourly prices to participants for changes in their usage from an established base amount; those hourly prices contain an explicit ($/kWh) marginal reliability (outage cost) element to reflect exigent reserve conditions (Barbose et al. 2004)

Market Impacts of Demand Response in Regions with Organized Wholesale Markets

About 60% of U.S. load is served by utilities or load serving entities that operate in regions with wholesale markets administered by ISOs/RTOs. Retail competition is also allowed in many of the states in these regions. These last-price wholesale electric commodity markets pay all competitively dispatched load a price determined by the last successful bid, which also sets the market clearing price. The market clearing price covers operating or production costs for the dispatched load (if each generator bids at least its marginal supply cost). If supply is very tight relative to demand, spot market energy prices will rise as more expensive units set the market clearing price. As a result, all units get the higher price, which includes creating "scarcity rents" for suppliers with costs below that of the marginal, price-setting unit.[80] Accordingly, spot energy prices serve as signals about whether additional supply- or demand-side capacity investments are needed, and what level of return to expect.

Three organized markets (NYISO, PJM, and ISO-NE) have established capacity payment mechanisms to create an additional stream of revenues for generators to recoup their investment costs. LSEs are required to purchase capacity in these markets to meet the expected peak demand of the customers they serve.

The impacts of demand response in an organized wholesale spot market are depicted in Figure B-5.[81]

The short-term market impacts of specific demand response events can be traced as follows:

- Depending on the timing and type of demand response option, customers' load changes may be integrated into day-ahead or real-time energy markets [as indicated by the arrows at the top of Figure B-5).
- Reductions in load during high-priced peak periods move marginal usage down the electricity supply curve (see Figure B-3), lowering market clearing prices during the demand response event (the event price in Figure B-5).
- This lowers LSEs' purchasing costs in the applicable wholesale market during the event. These savings may be captured by the LSE initially, but ultimately a significant share should be passed on to their customers (LSE event energy cost in Figure B-5).[82]

[80] This argument assumes that generators must recovery all of their revenue requirements and variable running costs, from energy sales at spot market prices. Some markets impose capacity requirements on LSEs that constitute a form of investment cost recovery for generators selling in those markets.
[81] The Midwest ISO (MISO), ERCOT and the California ISO (CAISO) all do not operate capacity markets.
[82] In some states, public utility commissions have adopted tariffs that specify the percent of savings that a regulated LSE providing default service must pass on to their customers. Eventually, competitive pressures should motivate LSEs to pass a significant portion of purchase cost savings to their customers.

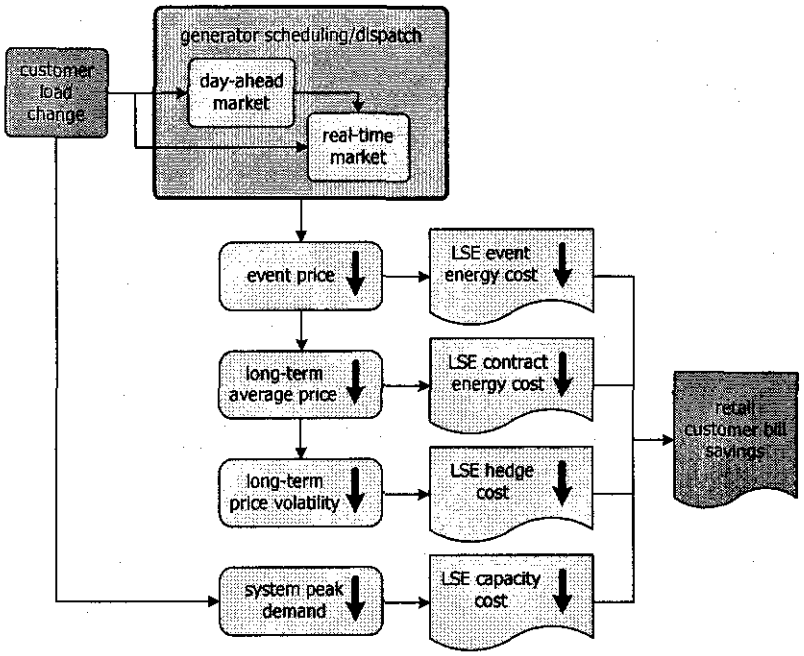

Figure B-5. Market Impacts of Demand Response in Regions with Organized Wholesale Markets

In regions with organized spot markets, demand response can produce cascading positive market impacts in the medium or long-term, realized over months or years (see Figure B-5):

- Reduced average market clearing prices can reduce forward contract costs for LSEs; these savings are then passed on to their customers (LSE contract energy cost in Figure B-5)
- Reduced volatility in market clearing prices puts downward pressure on risk premiums incorporated into hedged pricing products offered by competitive LSEs (LSE hedge cost in Figure B-5) and may lower transaction prices
- Lower forecast peak demand, resulting from demand response, also reduces LSEs' capacity acquisition requirements (LSE capacity cost in Figure B-5).

Long-term market impacts are less clear in organized wholesale and competitive retail markets compared to a vertically integrated utility system. A vertically integrated utility is allowed to directly pass through its capacity investment to customers in rates and likely most of its purchased energy and capacity costs as well; savings realized from demand response that avoids "uneconomic" investments or expenditures for peaking capacity are a direct source of cost savings to customers. In contrast, in organized spot markets, investment risk for new resources is assumed by the private sector. The combination of lower market clearing prices and reduced capacity requirements will dampen capacity investment signals, which should reduce construction of unneeded new power plants.

In summary, because organized spot markets use energy market clearing prices to pay generators for operating, but often only a fraction of the committed capacity costs, the long-term capacity savings benefits of demand response may not be fully monetized and

paid to demand response providers. Because the spot market valuation of demand response is linked to wholesale market clearing prices (for energy and capacity) rather than avoided capacity costs, this creates different payment streams and priorities between the two market structures. Policymakers need to recognize these differences in designing demand response options and evaluating benefits derived from market impacts under these different market structures.

Reliability Benefits

In addition to improving the efficiency of electricity markets, demand response can provide value in responding to system contingencies that compromise the dispatcher's ability to sustain system-level reliability, and increase the likelihood and extent of forced outages. Electric systems in the U.S. conduct long-term planning exercises to specify the level of resources required to serve the system's anticipated maximum load reliably in the long term. Typically, planning reserve margins are 15-18% of historic maximum system demand.

System operators arrange for some of the available generation resources to serve as reserves to cover real-time load-serving requirements and avoid outages; operating reserves of 5-7% of forecast demand must be maintained at all times. The system operator typically uses standby generators, ready to be run in less than 30 minutes, to deal with abrupt changes in load or unexpected loss of generator or transmission availability. Demand-response based load reductions can be used to replace some of this stand-by generation to rebalance load and supply.

Demand response can supplement system reliability by providing load curtailments that help restore reserves, providing incremental reliability benefits to the system.[83] Customers participating in emergency demand response programs receive incentive payments for reducing load when called upon by the system operator. They receive no up-front capacity payments in some program designs because they are not counted on as system resources for planning purposes. Instead, they are supplemental resources, the need for which is not foreseeable, or even likely, but possible. They represent an additional resource for reliability assurance, distinct from capacity-based demand response programs (see the textbox below).

[83] The capacity they provide can be particularly valuable if located in what operators call "load pockets", localized areas with a shortage of available resources to serve load when a generator is out of service.

Roles of Capacity and Emergency Demand Response Programs

Emergency demand response programs provide benefits distinct from capacity-based demand response programs. In capacity programs, customers are paid incentives based on the avoided cost of new generation capacity and are counted among planned reserves. As such, they become part of the overall portfolio of resources assembled to meet system reserve requirements. Capacity-based demand response does not provide incremental system reliability—it supplants conventional resources in meeting established reliability goals, simply replacing what a generator that was not built would have provided.

In contrast, emergency demand response programs provide incremental reliability benefits at times of unexpected shortfalls in reserves. When all available resources, including capacity demand response programs, have been deployed and reserve margins still cannot be maintained, curtailments under an emergency demand response program reduce the likelihood and extent of forced outages. Load curtailments under emergency demand response programs are therefore valued according to their impact on system reliability.[84]

System operators generally dispatch emergency demand response programs only after exhausting all available capacity and operating reserves. When operating reserves are called upon to go from standby status to actually producing energy to serve load, the level of remaining operating reserves drops if additional replacement resources are not available. This is analogous to a consumer drawing down savings to pay an unexpected bill, leaving them more vulnerable to consequences from further unanticipated expenses.

System operators can reduce this vulnerability by asking emergency program participants to curtail load, thereby reducing system demand and operating reserve requirements. This means that some generating resources can revert to their standby status and be ready for another contingency event, and can be likened to a cash infusion to restore savings in the consumer analogy. The curtailment allows the operator to maintain reliability at prescribed or target levels (Kueck et al. 2001). At the margin, this form of demand response provides value, although it is not priced in any market.

Figure B-6 illustrates this impact, and provides a way to estimate these reliability benefits. The portrayed system has been scheduled to provide D_1 units of energy (including required reserves) at a price of P_1 at a specific time.[85] As the delivery time approaches, a system contingency arises that effectively pushes the supply curve to the left (e.g., a generator outage) or customer demand to the right (e.g., an unexpected surge in demand, as portrayed in the figure by the move from D_1 to D_2), so that supply and demand no longer intersect. This reserve shortfall is represented by the demand curve D_2. Activating an incentive-based demand response program initiates customer demand reductions that bring system demand back to D_1, thereby eliminating the reserve shortfall.

[84] It is possible that an emergency demand response program, while not explicitly designed to fulfill capacity requirements , may nonetheless be capable of providing some level of capacity benefits as well.
[85] In this example, customer demand is represented by a vertical line, because in a reliability event, which occurs within minutes or seconds of power delivery, demand may be viewed as fixed.

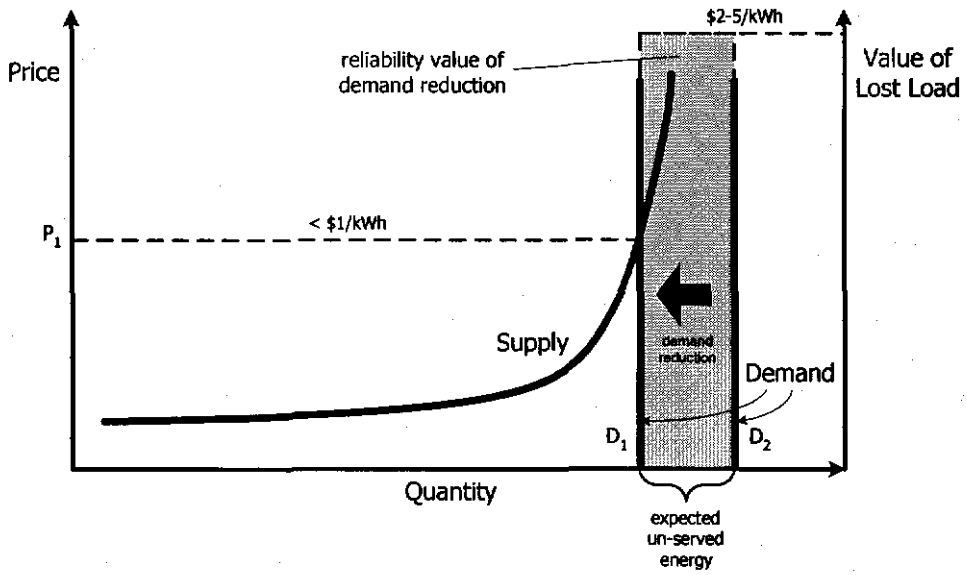

Figure B-6. Valuing the Reliability Benefits of Demand Response

While the price of served energy is determined by market conditions (P_1 in Figure B-6), the value of the demand reduction is defined by the decreased likelihood of a forced outage. Economists define the concept of *value of lost load* (VOLL) as the proper measure of improved reliability, since it reflects customer's marginal value for electricity under these circumstances. The product of VOLL and the *expected un-served energy* (EUE), the load that otherwise would not have been served, monetizes the value of the load curtailments (see the textbox below). This is represented by the shaded rectangle in Figure B-6 in the case where the curtailed load corresponds exactly to the amount of expected un-served energy.

Emergency demand response programs can provide low-cost, incremental resources to preserve reliability in various market structures; at present, the most prominent examples are implemented by the Northeast ISOs.

Value of Lost Load and Expected Un-Served Energy

"Value of lost load" (VOLL) is a measure of how customers value electric reliability, or what they would be willing to pay to avoid a loss of service. It varies among customers but is almost always greater than the retail price of electricity because customers incur costs from being disconnected without notice. Customer values factored into VOLL include inconvenience or discomfort, loss of sales or productivity (e.g., at retail premises or factories), large cleanup and restart costs (e.g., at pharmaceutical companies), and overtime costs to make up for lost production. Given the wide range of customer circumstances and difficulties in predicting which customers will be affected by a particular outage, the accepted industry practice is to adopt a VOLL of $2-5/kilowatt-hour (kWh), which represents an average value across the entire market.

"Expected un-served energy" (EUE) is a measure of the magnitude of a reserve shortfall. It takes into account the change in the likelihood of a curtailment and the consequences of such an event: how much load would have been forced off-line by dispatchers in such circumstances if the curtailments had not been undertaken. NYISO concluded that during the service restoration effort following the 2003 northeast blackout, demand response curtailments reduced forced outages kWh for kWh, because they enabled smoother service restoration. However, under other, less extreme conditions, curtailments were found to produce less than proportional reductions in EUE (NYISO 2003).

APPENDIX C. INTENSITY OF CUSTOMER DEMAND RESPONSE

This Appendix summarizes DOE's review of selected studies that have attempted to quantify the intensity of customer response to time-varying prices and demand response programs. First, different types of price elasticity used to measure demand response intensity are introduced. Next, the results of studies that estimated price elasticities for large and small customers exposed to time-varying rates are summarized. Some studies have examined the demand response intensity of programs targeting demand response-enabling technologies; these results are compared next. Finally, the results of studies that estimated load impacts from direct load control programs are summarized.

Indicators of Demand Response Intensity

For rate options and demand response programs that elicit load modifications directly in response to price changes, the intensity of customers' demand response is typically expressed in terms of their *price elasticity* (see the textbox below). Price elasticity provides a normalized measure of the intensity of customers' load changes in response to price circumstances. In analyzing price response, it is important to not confuse reported own-price and elasticity of substitution values. *Own-price elasticity* is defined as the percentage reduction in electricity usage in response to a one percent increase in the price of electricity. In analyzing price response among large industrial and commercial customers, it is common instead to estimate the *elasticity of substitution*, which measures the propensity of customers to shift electricity usage from peak to off-peak periods in response to changes in relative peak and off-peak prices. The substitution elasticity is defined as the percentage change in the ratio of peak to off-peak electricity usage in response to a one percent change in the ratio of off-peak to peak electricity prices. Various factors may influence customers' price elasticity, including the nominal level of prices. For example, some customers may be relatively unresponsive when prices are low but find it worthwhile to reduce load at very high prices. This characteristic of price elasticity has important implications for the design and evaluation of time-varying pricing and demand response programs.[86]

For DLC programs or other types of demand response programs where customers are not directly responding to a price, the intensity of customers' response is typically measured in terms of an absolute or relative load impact (e.g., kW or percent load reduction).

[86] If price response increases with relative prices, then it is important to account for this factor when estimating how customers will respond to prices or to a demand response program incentive. A specific price threshold may be necessary to obtain a significant response among a group of customers.

Price Elasticity: Insights and Sources of Confusion

Price elasticity is a normalized (for the relative price change) measure of the intensity of how usage of a good (in this case electricity) changes when its price changes by one percent. It facilitates a comparison of the intensity of load changes among customers since the price change has been factored out; the price elasticity is a relative measure of response. For example, Customer A, with an elasticity of 0.25, responds to the same relative price change much more than Customer B, who has an elasticity of 0.05 (i.e., five times more relative to the customer's usage level). But, not five times greater than another customer in absolute terms, unless they have exactly the same load. This highlights the relative comparison of intensity that a price elasticity response provides; the basis is each customer's load. Consequently, some studies prefer to report and compare customers' actual percentage changes in load. This is insightful, as long as the load changes were in response to the same change in prices.

A potential source of confusion comes from differences in how price elasticity is reported. Some analysts report the *own-price elasticity*, which is expected to be negative, since a one percent increase in price would be expected to cause usage to go down, all other things equal. It is a useful measure of how customers adjust to increases in the price of electricity by adjusting the consumption of other goods. This is especially useful when evaluating longer-term adjustments to changes in electricity price. Other analysts report the *substitution elasticity*, which takes on only positive values. The substitution elasticity focuses on how consumers substitute one good for another, or goods in different time periods for one another, when relative prices change. Specifically, if the price of electricity varies substantially from one time period to another, and customers can shift usage among those periods, then the appropriate measure of price response is how relative usage changes in those periods. The substitution elasticity is therefore defined as the relative change in usage in the two periods (e.g., the ratio of the peak to off-peak usage) for a one percent change in the relative prices in those periods (the ratio of the off-peak to peak price). Note that the price term uses the inverse price ratio, which is why substitution elasticities are positive (e.g., a higher peak price decreases the off-peak to peak price ratio, causing peak load to be reduced and therefore the peak to off-peak load ratio to decline).

On an absolute value basis, ignoring the sign, own-price and substitution elasticities are similar in that they both measure relative changes, so a value of zero corresponds to no change in usage regardless of the change in price (i.e., perfectly price inelastic), and absolute values progressively greater than zero indicate relatively higher price response. They are roughly similar measures of intensity on a nominal basis—a substitution or an own-price elasticity of 0.50 both indicate relatively high changes in load in response to price changes. But because these two elasticity values measure a different characterization of how usage is adjusted to price changes (i.e., price in one period vs. relative prices in two periods), there is no simple way to cross-map reported values. They should be used in the appropriate context: the own-price elasticity when the circumstances involve reduced electricity usage and the substitution when shifting from one time to another characterizes price response.

In this report, substitution elasticities are always reported as a positive number and own-price elasticities as a negative number.

Price Elasticity Estimates

For mass-market (residential and small commercial) customers, there is an extensive price elasticity literature examining the load impacts from TOU rates. Not surprisingly, the estimates produced by these various studies span a wide range, reflecting both methodological differences and situational factors (e.g., related to customer

characteristics or program design). Caves et al. (1984) pooled data from five residential TOU pilots implemented in the U.S. in the latter half of the 1970s (see Table C-1). The average elasticity of substitution derived from this pooled data set was 0.14, but elasticities varied by a factor of three, from 0.07 to 0.21, depending on the household's electric appliance holdings (Faruqui and George 2002). King and Chatterjee (2003) reviewed price elasticity estimates from 35 studies of residential and small commercial customers published between 1980 and 2003. They report an average own-price elasticity of –0.3 among this group of studies, with most studies ranging between –0.1 and –0.4. Several studies have also examined the intensity of residential (and small business) customers' response to CPP and RTP tariffs and isolated the affect of various factors and customer circumstances. A recent study at Commonwealth Edison in Illinois of the first residential RTP pilot in the U.S. found notably lower demand response intensity than has been observed for small customers; own-price elasticities were –0.04 in 2003 and –0.08 in 2004 (Summit Blue Consulting 2005). However, the weather during these two summers was unseasonably cool and A/C usage and hourly prices were correspondingly low, which suggests that the price response may be higher under more extreme conditions.

An evaluation of a recent residential CPP pilot in California estimated a statewide average elasticity of substitution of 0.09 on critical peak days occurring between July and September and reported that the average statewide reduction in peak period energy use on critical peak days was about 13% (Faruqui and George 2005).[87] However, the elasticity varied by more than a factor of three across five climate zones, reflecting regional trends in temperature and A/C saturation (which varies from 7% to 73% of households). The study also found substantial differences between customers' price elasticities during the hotter summer months (July—September) and during the shoulder months of May, June and October—also indicative of differences in A/C usage.

Information on the price elasticity of large commercial and industrial (C&I) customers is based primarily on studies that examined customers' response to RTP. These studies have employed several types of demand models producing different types of price elasticity measures and have examined variations with time of day, price level, and customer characteristics (e.g., business type, presence of onsite generation, number of years on RTP).

[87] Impacts varied across climate zones, from 7.6% in the relatively cool coastal climate zone (e.g. which includes San Francisco) to 15.8% in inland, hot climates of California (Faruqui and George 2005).

Table C-1. Demand Response Program and Pricing Studies: Estimated Price Elasticity of Demand

Type of Program	Target Market	Region (Utility)	Demand Response Impact (average per customer)	Comments
TOU	Residential	U.S (utilities in five states)	Elasticity of Substitution **0.14** average; **0.07** to **0.21** range depending on electric appliance holdings	Pooled results from five residential TOU pilots in the late 1970s. Sources: Caves *et al.* (1984) and Faruqui and George (2002).
TOU/ CPP	Residential and Small Commercial	U.S. and International (various utilities)	Own-Price Elasticity **-0.3** (average of 35 studies); **-0.1** to **-0.8** range across the studies	The authors calculated the simple average of own-price elasticity estimates from 35 studies of TOU or CPP. Source: King and Chatterjee (2003)
CPP	Residential	California (PGE, SCE, SDG&E)	Elasticity of Substitution **0.09** average (July-Sept.); **0.04** to **0.13** range across climate zones	Population of about 1,000 residential customers, including control groups, in 2003/4 California Statewide Pricing Pilot. Elasticity range across climate zones attributed to differences in A/C saturation (7-73%). Source: Charles River Associates (2005)
Day ahead RTP	Residential	Illinois (Com Ed, Community Energy Cooperative)	Own-Price Elasticity **-0.04** average (2003); **-0.08** average (2004); **-0.05** to **-0.12** range across customer segments (2004).	Population of about 1,000 customers in 2004; $0.12/kWh maximum hourly price. Own-price elasticities were reported for six different customer segments defined in terms of housing type (single- or multi-family) and A/C equipment type (window, central, or none). Source: Summit Blue Consulting (2005)
	Med./Large C&I (>200 kW)	Georgia (Georgia Power)	Own-Price Elasticity **-0.01** to **-0.28** range across customer segments and hourly price levels	Population of about 1,600 customers. Elasticities were estimated for seven different customer segments at four different price levels, ranging from $0.15 to $1.00/kWh. Source: Braithwait and O'Sheasy (2002)
	Med./Large C&I (>100 kW)	U.K. (Midlands Electric)	Hourly Own-Price Elasticity **-0.01** to **-0.27** range in maximum hourly elasticities, across customer segments	Population of about 500 customers, most with peak demand >1 MW. Hourly own-price and substitution elasticities were calculated for each of five different industry classifications. Source: Patrick and Wolak (2001)
	Large C&I (>1 MW)	North and South Carolina (Duke Power)	Average Peak-Period Own-Price Elasticity **< -0.01** to **-0.38** range across customers	Population of about 50 customers, some with 8 years experience on RTP. Hourly own-price were calculated for each customer, and averaged over the peak period (2:00-9:00 p.m.). Source: Taylor *et al.* (2005)
	Large C&I (>1 MW)	Southwest U.S. (Central and Southwest Services)	Elasticity of Substitution **0.10 to 0.27** range across customer segments and definitions of the peak period	Population of 54 customers, segmented into two groups, with firm day-ahead hour-ahead notice of hourly prices. Elasticities estimated for each group and for different definitions of the peak period. Source: Boisvert *et al.* (2004)
	Large C&I (>2 MW)	New York (Niagara Mohawk)	Elasticity of Substitution **0.11** (average); **0.02** to **0.16** range across customer segments	Population of about 150 customers. Individual customer elasticities vary substantially within sectors: e.g., most manufacturing customers are either highly responsive or not at all. Source: Goldman *et al.* (2005)

Note: Elasticity values are the averages of all participants' elasticity at all price levels, unless otherwise noted. Elasticity of substitution values are for intraday substitution between peak and off-peak periods, while own-price elasticities are the average value, unless noted as hourly.

Braithwait and O'Sheasy (2002) analyzed data from participants in Georgia Power's RTP program, the largest in the country. The authors estimated own-price elasticities for seven

different business customer segments and examined differences across hourly price levels. Most customer segments exhibited larger price elasticities at higher prices. The most responsive customer segment was a group of very large industrial customers (peak demand > 5 MW) who, in exchange for slightly lower base rates, had opted to receive notification of hourly prices on an hour-ahead (rather than day-ahead) basis. This group exhibited a price elasticity of –0.18 to –0.28 across the range of reported prices ($0.15/kWh to $1.00/kWh), which was double the elasticity of any other group. The least responsive customer segments, consisting of smaller C&I customers that neither had onsite generation nor had previously participated in the utility's curtailable rate, exhibited price elasticities of –0.06 or lower at all price levels.

A study of about 150 large customers at Niagara Mohawk estimated an average substitution elasticity of 0.11 among those that faced day-ahead hourly prices (Goldman et al. 2005). However, the average elasticity varied substantially across business categories (e.g., average elasticities were 0.16 for manufacturing customers, 0.10 for government/education customers, and 0.02 for health care facilities) and even more within them (e.g., half of the industrial customers were very inelastic, and half were relatively elastic).

Studies of the large C&I RTP programs offered by Duke Power and Midlands Electric (in the U.K.) estimated average hourly own-price and substitution elasticities (Taylor et al. 2005, Patrick and Wolak 2001). Both studies found a substantial range in own-price elasticity values over the course of the day and among customers. Among the 50 or so participants in Duke's program, the average hourly price elasticity during peak period hours ranged from less than –0.01 to –0.38. This study also concluded that many large C&I customers exhibit complementary electricity usage across blocks of afternoon hours. That is, high prices in one hour result in reduced usage in that hour as well as in adjacent hours. This is consistent with industrial batch process loads that, once started, must continue for a specified period, and with other business practices that exhibit similar relationships (e.g., rescheduling of labor shifts). Usage in many other hours of the day was found to be a substitute to the afternoon hours. The study of Midlands Electric's customers also found substantial variation in the magnitude and hourly pattern of price elasticity among different industrial classifications. Customers in the water supply industry were the most price-responsive, with a maximum hourly own-price elasticity of –0.27, while all of the other industrial classifications in the participant population exhibited price elasticities of less than –0.05 in all hours.

Impact of Enabling Technologies on Price Response

A small number of utilities have offered pilot programs targeted at mass market customers that integrate CPP with enabling technology, specifically load control devices that receive price signals and can be programmed by customers to reduce A/C or other loads during critical peak periods (see Table C-2).

Table C-2. Load Response from Enabling Technologies in Combination with CPP

Enabling Technology	Target Market	Region (Utility)	Demand Response Impact (average per customer)	Comments
Thermostat reset	Residential	California (SDG&E)	**0.64 kW (27%)** average peak period load reduction on critical peak days; **0.4 kW** attributed to enabling technology.	2003/2004 pilot program with about 220 residential customers and about 235 C&I customers, including control groups. Customers had "smart thermostats" that could be programmed to raise the temperature set point during critical peak periods. Analysis distinguished between enabling technology and behavioral components of price response. Peak period prices on critical peak days averaged $0.65/kWh for residential customers, $0.87/kWh for customers with <20 kW peak demand and $0.71/kWh for larger C&I customers. Source: Charles River Associates (2005)
	Small/Med. C&I (<200 kW)	California (SCE)	Customers with <20 kW peak demand: **0.95 kW (14%)** average peak period load reduction on critical peak days; attributed entirely to enabling technology. Customers with 20-200 kW peak demand: **3.1 kW (14%)** average peak period load reduction on critical peak days; **2.5 kW** attributed to enabling technology.	
Control of multiple loads (A/C, heat pump, water heater, pool pump, and/or appliances)	Residential	New Jersey (GPU)	Elasticity of Substitution **0.3** (average)	Pilot program results from summer 1997. Critical peak price was $0.50/kWh. Source: Braithwait (2000)
	Residential	Florida (Gulf Power)	**2.7 kW (41%)** average load reduction during critical peak periods	Estimated response from current *GoodCents Select* program. Source: Borenstein *et al.* (2002).
	Residential	Upper Midwest (AEP)	Winter: **3.5-6.6 kW** Summer: **1.5-2.0 kW**	Pilots conducted at three AEP utilities in the early 1990s with about 600 customers, including control groups. Critical peak price ranged from $0.15-$0.29/kWh among the three utilities. Source: Levy Associates (1994)

An evaluation of the recent Statewide Pricing Pilot in California sought to quantify the incremental impact of this type of technology on customers' demand response. Groups of residential and small commercial participants in this pilot faced CPP and had "smart thermostats," which customers could pre-program to automatically raise their temperature settings by a specified number of degrees during critical peak periods. The statistical model used in the evaluation decomposed these customers' total load reduction during critical peak periods into a "technology component" (i.e., the portion of the load reduction attributable to use of the smart thermostat) and a "price component" (i.e., the portion attributable to manually-implemented actions). The average load reduction by residential customers with smart thermostats during critical peak days was approximately 0.64 kW, approximately two-thirds of which was attributed to use of the smart thermostat. Among small business customers, the relative impact of the enabling technology was even more pronounced.

A handful of utilities elsewhere in the U.S. have implemented residential CPP pilots in which participants were provided with thermostats that they could program to control their A/C and other appliances (pool pumps, heat pumps, and electric water heaters)

during critical peak periods. Studies of these programs have typically found that participants exhibited a relatively high intensity of demand response. For example, an analysis of GPU's pilot (in New Jersey) measured a substitution elasticity of 0.3, which is higher than most elasticity of substitution values estimated from residential TOU pilots (Braithwait 2000). Studies at Gulf Power and American Electric Power (AEP) where multiple loads could be controlled in response to critical peak prices reported that average load reductions among a sample of customers were in the 35-40% range (Levy Associates 1994).

Load Impacts from Direct Load Control

Approximately 180 U.S. utilities (out of the 1,118 investor-owned, municipal, and rural cooperative utilities that reported demand-side management efforts) report that they currently offer residential DLC programs that primarily target specific appliances, such as air conditioners or water heaters, of mass market customers (EIA 2004).[88] Various control strategies (e.g., cycling the device on and off at a specified frequency, shutting the device off, or resetting a thermostat set-point) are utilized during prescribed conditions depending on end use, control equipment vintage, and program design.[89] Several of these programs have conducted relatively recent measurement and evaluation studies with results that are publicly available. In DLC programs, because the utility controls the switch, the customer cannot be said to exhibit price response, per se, although the change in the customer's load is measurable. The most appropriate measure of demand response impact for this program type is simply the average or expected load reduction (in absolute or percentage terms), rather than the price elasticity.

Table C-3 summarizes the measured impact from selected evaluations of DLC programs that targeted customers with air conditioning or water heating load control devices. The results indicate the range of possible load impacts, although the individual values are not readily comparable because of the differences in program design features, cycling strategies, and climate. DLC programs targeting residential A/C have reported load reductions ranging from approximately 0.4 to 1.5 kW per customer over the course of an event. The magnitude of the load reduction per customer can strongly depend on climate, the corresponding level of A/C usage that would occur absent load control, and the control strategy deployed (e.g. 100% shed, duty cycling). Furthermore, when customers have the ability to over-ride the curtailment via their thermostat, the average response per customer has generally been found to decline (sometimes substantially) over the course of each event. Residential water heating DLC programs have yielded load reductions in the range of 0.2 to 0.6 kW per house. The magnitude and timing of the load impact depends on equipment size, ground water temperature and household size and operating use patterns.

---

[88] Demand-side management efforts include energy efficiency and/or load management programs.

[89] In newer DLC programs, particularly those that use thermostat-based controls, customers can typically over-ride curtailments on an event-by-event basis, either by pushing an "over-ride" button on their thermostat, logging onto a program website, or calling the utility. If they do over-ride a curtailment event, customers typically forfeit a portion of their incentive payment or are charged a penalty.

Table C-3. Direct Load Control Programs: Estimated Load Impacts

Type of Program	Target Market	Region (Utility)	Demand Response Impact (average per customer)	Comments
A/C temp. reset (with over-ride option)	Residential	SDG&E	**0.44 kW** (average); **0.10-0.81 kW** (range over 12 events)	Sample of about 100 customers (including control group) with 12 test events in summer 2004. Source: KEMA-Xenergy (2004)
A/C cycling (with over-ride option)	Residential and Small Commercial	New York (LIPA)	**0.75-0.91 kW** (residential) **1.01-1.43 kW** (small commercial)	Ranges in average hourly load reductions over a single event day with 50% cycling. Based on 12,000 residential customers and 2,000 commercial customers. Source: Lopes (2004)
A/C cycling (no over-ride option)	Residential	Minnesota (Xcel Energy)	**1.27 kW**	Based on interval metering at large number of customer sites; 50% cycling frequency. Source: Xcel Energy (2004)
		California (SMUD)	**0.71-1.59 kW**	Pilot program results from summer 2002. The lower bound corresponds to a cycling frequency of 33% and outdoor temperature of 96-100° F; the upper bound corresponds to a cycling frequency of 66% and an outdoor temperature of >100° F. Source: Violette and Ozog (2003).
		Kentucky (LG&E, KU)	**0.52-1.12 kW**	Interval metering measurements at 20 customer sites. The lower bound corresponds to a cycling frequency of 33% and outdoor temperature of 90-95° F; the upper bound corresponds to a cycling frequency of 66% and an outdoor temperature of >95° F. Source: Violette and Ozog (2003).
		Maryland and D.C. (Pepco)	**0.96 kW** (MD) **0.76 kW** (DC)	Measured impact for hour ending 17:00, based on 20-year average system peak day weather; 43% cycling off strategy. Source: Horowitz (2002)
Electric water heater cycling		Oregon (PGE)	**0.65 kW**	Load reductions measured at 0800. Source: PGE (2004)
		Maryland (BGE)	**0.2 kW** (at 5 PM) **0.3 kW** (at 7 PM)	Load reductions measured at 1700 and 1900. Source: BGE (2002, 2003)

APPENDIX D. STANDARDS, PROTOCOLS AND PRACTICES FOR ESTIMATING THE BENEFITS OF DEMAND RESPONSE

In Section 4 of this report, DOE offers several recommendations on establishing standardized methods and protocols and enhancing practices for estimating the benefits of demand response. This Appendix provides further discussion that supports these recommendations.

1. DOE recommends that stakeholders collaborate to adopt conventions and protocols for estimating the benefits of demand response and, where appropriate, develop standardized tests that evaluate demand response program potential and performance.

Policymakers and industry participants should develop standardized tests that are applicable and appropriate for the evaluation and cost-effectiveness screening of demand response resources. Standard Practice Manual (SPM) tests are widely used among state regulatory commissions and utilities to evaluate and screen energy efficiency programs (CPUC 2001).[90] Historically, a number of states and utilities have also used these tests for cost-effectiveness screening of load management programs and, recently, there have been some efforts to modify the SPM tests to enhance their usefulness for evaluating demand response resources in the context of competitive wholesale markets (CPUC 2003; Violette et al. 2006, Orans et al. 2004). However, there is general consensus that a more comprehensive evaluation framework is needed to fully capture the benefits of demand response (PIER DRRC, 2005).

Some of the challenges in developing standardized tests appropriate for demand response are revealed by comparing energy efficiency and demand response resources. While it is relatively straightforward to identify and estimate the peak demand and energy reduction impacts of energy efficiency, this is much more difficult for most demand response options. Because most demand response options are relatively new, our ability to predict program participation rates and assess how specific program designs and dynamic pricing affect customer behavior is still rudimentary.[91] Moreover, many forms of demand response turn on behaviors that are price- or incentive-driven, and may change in response to changing market circumstances. Uncertainties in estimating demand response impacts over a multi-year period mean that demand response benefit (and cost) estimates are equally uncertain.

---

[90] The SPM describes several tests that evaluating demand-side management programs from various perspectives: Participant Test, Ratepayer Impact Measure (RIM) Test, Total Resource Cost (TRC) Test, and Program Administrator (formerly Utility) Test.

[91] Load reduction impacts are well characterized for residential DLC programs that have operated for many years, although there have been issues in determining the extent to which customers remove load control switches or over-ride load curtailments. For interruptible/curtailable programs, little information exists from which long-term performance can be predicted. For thermostat-based programs, limited information gathered through several large pilots is available to shed light on customer behavior. For optional RTP tariffs, substantial evidence shows that customer attrition can be a significant problem when major price shocks occur.

In contrast, 15-20 years of implementation experience and tens of millions of dollars spent evaluating energy efficiency programs has produced well-developed methods for forecasting market penetration and estimating first-year energy savings, expected economic lifetime and the persistence of savings for most energy-efficiency measures and programs. This task is further eased because most energy efficiency measures produce savings that are not dependent upon customer behavior.

The SPM tests, which use avoided costs to characterize benefits, have shortcomings in the way in which they characterize the value of demand response to the electric system and customers. Despite recent advances, these tests are not well suited to estimating the value of dispatchable demand response resources. For example, SPM tests have limited ability to reflect the value of capacity in critical peak hours, and the potential of demand response to mitigate episodic, high spot market prices is therefore undervalued. Other aspects of demand response benefits, such as quick ramp-up (relative to constructing new generation resources), and reliability benefits, are also not captured by SPM tests. A more comprehensive analytic framework is needed to fully evaluate and assess the benefits of demand response. At present, summarizing the benefits and costs for some types of demand response resources by means of a standardized test may be premature.

2. DOE recommends that these protocols: (1) clarify the relationships and potential overlap among categories of benefits attributed to demand response to minimize double counting, (2) quantify various types of benefits to the extent possible, and (3) establish qualitative or ranking indices for benefits that are found to be too difficult to quantify.

Policymakers and analysts assessing the merits of demand response mechanisms need to clarify the relative importance of benefits that are difficult to quantify.

Some demand response advocates allude to benefits, such as market power deterrence, risk mitigation and avoided pollutant emissions—that are not quantified but are presumed to be substantial (PLMA 2002; NEDRI 2003; Violette et al. 2006).[92] Not only are such benefits difficult to quantify, but care must be taken to avoid double-counting benefits from other sources (e.g., market-power reduction benefits must be disentangled from other market price impacts). Parties seeking to justify greater expenditures on demand response often assert the existence of such benefits. Policymakers, however, are often wary of including these benefits as criteria for designing policies to foster demand response. Research to determine the magnitude of these impacts and to develop methods for quantifying or incorporating them into benefit/cost analyses, without double counting, is needed.

3. DOE recommends that FERC and state regulatory agencies work with interested ISOs/RTOs, utilities, other market participants, and customer groups to examine

[92] These non-quantified demand-response benefits are discussed in more detail in section 3 (see *Other Benefits*).

how much demand response is needed to improve the efficiency and reliability of wholesale and retail markets.

It is appropriate for state and regional policymakers to ask how much demand response is sufficient for their specific market structure and system conditions. A number of demand response studies confirm that a little demand response can go a long way towards improving the efficiency and operations of electricity markets, both in theory and practice. However, existing studies do not address how to identify optimal, or target, levels of demand response in specific market settings. Initiatives should be launched at the appropriate market level (e.g. state or region) to establish relevant goals and appropriate targets for demand response.

As part of the process of determining how much demand response is needed, it is also important to address the appropriate mix of different types of demand response options (e.g. emergency demand response programs, direct load control, time-varying pricing) and any timing issues related to demand response resource deployment and ramp-up (Violette et al. 2006). Although this is not a problem today given the low participation rates in dynamic pricing and demand response programs, it is important to acknowledge that there may be a potential for diminishing returns in the value of demand response beyond certain levels of saturation. For example, the level of price-based demand response is somewhat self-limiting—if at some point demand response becomes widespread, customers may find that their savings from load response actions deteriorate as the impact of their collective response on market prices grows.

4. DOE recommends that regional planning initiatives examine how demand response resources are characterized in supply planning models and how the benefits are quantified. More accurate characterization of certain types of demand response resources may require modifications to existing models or development of new tools.

Resource planning methods currently used to characterize demand response resources are too constraining and rigid to capture the full benefits of all types of demand response resources. In vertically integrated systems, long-term resource planning models characterize demand response as a way to avoid generation (and in some cases transmission and distribution) investment costs. Demand response is typically portrayed as a generation unit, which can either be dispatched indiscriminately or with some restrictions on the total frequency or hours of service. This characterization does not fully describe the differences between generation and demand response resources.

Certain types of demand response resources provide benefits that generation cannot. For example, capacity-based demand response programs can provide equivalent capacity to generation investments but with greater flexibility. This is because some types of demand response resources can be implemented more quickly than a power plant can be sited and built, and customers often prefer or are willing to accept a shorter time commitment than

is necessary to amortize a power plant.[93] These flexibility benefits are particularly important from a system cost perspective that includes and explicitly accounts for the uncertainties in demand growth or generation unit retirement schedules and costs. Resource planners' avoided cost studies should explore the implications and value of flexible demand response program options as both long-term and short-term operational resources to deal with generation load balance and transmission and distribution adequacy challenges.

Moreover, long-term resource planning models often do not fully recognize or represent the benefits of price-based options such as RTP. RTP ties hourly retail prices to prevailing wholesale market supply costs. To fully account for its potential benefits, RTP should be portrayed as a change in demand in response to prices, not as a resource dispatched to serve demand. Moreover, the RTP prices in tariffs offered by vertically integrated utilities often reflect both marginal supply costs and reliability value of load curtailments. These hour-by-hour impacts, which are carefully measured in ISO/RTO demand response program performance studies, can get overlooked in a long-term resource planning exercise.[94]

On the other hand, peaking generation resources have some characteristics that are more desirable to resource planners than demand response resources. For example, system operators have high confidence that generation resources will come online when needed, whereas customers may decide not to respond when a demand response resource is called. This makes it more difficult to predict the precise amount of available resources on a given day. Another advantage of supply resources is that they can provide certain ancillary services, such as voltage support and re-starting the electrical grid after a blackout, that demand response resources cannot. These considerations should also be incorporated into planning models to appropriately characterize and assess available resources.

5. DOE recommends that, in regions with organized wholesale markets, ISOs and RTOs should work with regional state committees to undertake studies that characterize the benefits of demand response under foreseeable future circumstances as part of their regional transmission expansion plans as well as under current market conditions in their demand response program performance studies.

---

[93] The capacity programs implemented by several ISOs do not involve long-term customer commitments (customers may participate for only a few months if they wish). These programs have demonstrated reasonably predictable and stable performance without putting "iron in the ground"—generation assets whose costs must be recovered over 20 years or more (NYISO 2003). Emergency programs that require no commitment on the customer's part have attracted substantial participation by customers that delivered curtailments on a pay-for-performance basis, and are a potentially cost-effective way to increase system reliability.

[94] Moreover, RTP may result in increased usage during off-peak periods when prices are lower. Increased unit utilization lowers the overall average cost of capital, another important source of benefits that may not be adequately reflected in current study practices.

In regions with organized spot markets, analytic methods focus primarily on assessing the short-term impacts of ISO/RTO demand response programs; more work is needed to assess the potential long-term benefits of demand response resources. ISOs/RTOs that offer demand response programs provide annual performance assessments to FERC that focus primarily on realized, short-term impacts. These assessments provide policymakers, market participants, and customers with information on both the level and distribution of demand response benefits and resource costs.[95] However, in the absence of a forward electricity market that would create a steady stream of guaranteed annual benefits, the value of demand response necessarily depends primarily on current market conditions.

However, ISOs and RTOs can and should provide information on the future value of demand response within their regional markets. Most ISOs and RTOs conduct or coordinate long-range planning studies that focus on developing coordinated system expansion plans that identify projects that can ensure electric system reliability, reduce congestion and also provide market signals for planning and running generation and transmission systems and demand-side management projects (ISO-NE 2005b; PJM Interconnection 2005b). One goal of the studies is to use forecasts of regional load/resource balance to identify needed investments to forestall potential supply shortfalls that could lead to high price volatility. The extent to which demand response is considered in these regional transmission expansion plans is evolving over time. ISOs, RTOs and regional state committees are well positioned to recognize the long-term benefits of demand response and incorporate demand response into their long-term system plans.[96] Another option would be to facilitate a forward market in demand response, as PJM has proposed (PJM Interconnection 2005c).

[95] Because benefits can vary from year to year and opportunities to participate are not always available, it is important that load aggregators and customers are made aware of how benefits and costs (i.e., incentive payments) may vary with market circumstances.

[96] Efforts are already beginning in this area. A recent pilot study by ISO-NE that compared the value of RTP and other types of demand response programs under alternative market circumstances was intended to facilitate discussions of this issue among policymakers, ISOs, load serving entities, and customer groups (Neenan Associates 2005).