The Role of Energy Efficiency and Energy Service Companies in Wholesale Markets

Beth W. Dunlop, Grueneich Resource Advocates
Donald Gilligan, Predicate Energy, LLC

ABSTRACT

This paper examines the potential for energy efficient technologies and practices to reach beyond the retail market into the wholesale energy market. Energy service companies (ESCOs) successfully deliver energy efficiency into the retail market. The wholesale electricity market, with billions of dollars of transactions, provides new opportunities to deliver energy efficiency. If this potential is realized, wholesale demand reductions can benefit customers and others by: 1) reducing market prices through a more elastic demand curve, 2) increasing market liquidity, 3) reducing price volatility, 4) moderating the effects of supplier market power, 5) delaying the need for new power supplies, and 6) increasing system reliability.

This paper first addresses opportunities for energy efficiency in wholesale electric markets and describes the different uses of permanent and temporary load reductions. Next, the paper identifies the several benefits that demand responsiveness provides to wholesale markets, including mitigation of market power and price volatility, increasing liquidity in the market, delaying the need to construct new supply, and overall price reductions. Third, several of the first demand responsiveness programs are described. Finally, the paper compares and contrasts the information and technologies used to deliver retail versus wholesale energy efficiency.

Introduction

The purpose of this paper is to introduce the concept of demand responsiveness, describe how and why this concept may open new opportunities for energy efficiency and energy service companies (ESCOs), and report on progress with actual demand responsiveness programs.

One of the major questions that has not yet been fully addressed is the magnitude of the market potential for energy service companies, especially for temporary load reductions. Market potential will depend on a number of factors, including the following:

- ESCOs' ability to market the concept to customers.
- Cost and ease of installation and operation of load curtailment equipment.
- Cost of telemetry, metering, and communications equipment.
- Design of ISO, power pool and utility demand responsiveness programs, especially their price signals, program complexity, and telemetry requirements.

Wholesale energy markets and energy efficiency existed in separate worlds until the recent restructuring of electricity markets and increasing capacity shortages in some states. Combining these worlds adds value to both. Wholesale markets improve their efficiency by strengthening the demand side of the equation, and energy efficiency concepts, products and

services begin to access even greater portions of their technical and economic potential through their value to wholesale markets. Synergies between retail and wholesale demand reductions, and between permanent and temporary load reductions, will make more efficiency projects cost effective.

Greater reliance on energy efficiency in the wholesale markets can also enhance grid reliability. In California, certain areas of the grid are heavily constrained during some time periods. The traditional solutions of building more transmission or new generation are expensive, time-consuming, and face considerable NIMBY ("Not In My Backyard") pressures. The CAISO and utilities are expanding the possible solutions to include energy efficiency and distributed generation. One example is the Tri Valley Request for Proposals, issued by the CAISO on January 18, 2000, which sought non-wires, i.e., generation or load management, alternatives to the local utility's proposed transmission upgrade.[1]

In the near term, load management is also valuable to the grid as the CAISO attempts to mitigate intrazonal congestion. Load is eligible to submit adjustment bids to the CAISO; in other words, a load submits a set of price/quantity bids to the CAISO. If needed, load on the congested side of a tie curtails and receives the same price that a supply resource would receive for providing the same service.

This paper first addresses opportunities for energy efficiency in wholesale electric markets and describes the different uses of permanent and temporary load reductions. Next, the paper identifies the several benefits that demand responsiveness provides to wholesale markets, including mitigation of market power and price volatility, increasing liquidity in the market, delaying the need to construct new supply, and overall price reductions. Third, several of the first demand responsiveness programs are described. Finally, the paper compares and contrasts the information and technologies used to deliver retail versus wholesale energy efficiency.

Opportunities for Energy Efficiency in Wholesale Markets

In the retail markets, the role of ESCOs is to provide energy efficiency products and services. Energy efficiency is viewed as a product or service separate from the fundamental supply/demand equation. For institutional reasons (e.g., rate freezes, rate schedules whose prices are averaged over long time periods, bundled rates, monopoly utilities as default providers, a history of utilities building supply to meet inelastic demand curves), and because of the scale issue, demand curves are more inelastic than supply curves in the retail market. In theory, customer choice and Energy Service Providers (ESPs) will expand over the long run, and after the rate freeze ends and the market is more open, ESPs will have more incentive to bid more elastic demand curves. Eventually, energy efficiency can participate in both the wholesale and retail markets in a way that is more tightly connected to the demand curve, rather than being essentially a separate product or service.

The problem that ESCOs solve is scale. Individual loads are too small to participate in the PX or ISO markets, wouldn't have an impact individually anyway, and the metering, bidding, settlements and billing infrastructure is too expensive for individual small loads.

[1] Although the Tri-Valley process considered non-wires options, on April 27, 2000, the CAISO Governing Board directed PG&E to proceed with its transmission expansion.

What ESCOs provide, at a minimum, is the technology to do load control for permanent or temporary reductions. In addition, ESCOs can aggregate customers into large enough blocks that power pools and ISOs are willing to deal with them, and manage or contract out the bidding, metering, settlements and billing processes. The ESCO strength is applying a comprehensive set of technology solutions at a customer facility. The next step for ESCOs is to expand the suite of technologies installed to include the equipment needed to make an impact in the wholesale market.

Wholesale markets represent an entirely new realm for energy efficiency (Kirby and Hirst 1999). Although substantial potential for additional cost-effective energy efficiency exists in the retail markets (California alone anticipates saving approximately 600 million kWh through its 2000 public-goods-charge-funded energy efficiency programs (Carter 2000)), it is the price signals of the wholesale markets, newly opening to demand-side bidding, that provide the potential for a true market-based opportunity for demand and energy reductions and the accompanying environmental benefits. The absence of useful price signals to end-use customers has become one of the most oft-lamented market barriers to greater penetration of energy efficiency. While electric industry restructuring has not yet advanced to the stage where every end-use customer sees and has the means to respond to price signals in real time, California and a few other states have begun to recognize that restructuring in an era of capacity shortage[2] is not going to work well unless that fundamental market barrier is erased or at least diminished.

Absence of accurate price signals is but one market barrier to the full penetration of all cost-effective energy efficiency. The existence of other market barriers (e.g., lack of information, perceived risk of energy efficiency investments, misplaced incentives (Hirst 1991)) and the remaining cost-effective potential for energy efficiency more than justify the continuation of existing retail strategies, including efficiency programs funded through public goods charges. However, the initial crumbling of one major market barrier is a breakthrough: the dryness of phrases like “more elastic demand curves” camouflage the significance of customers having access to price signals, albeit a little indirectly through the wholesale markets.

The enormous promise for energy efficiency and load management in restructured, capacity-constrained markets such as California’s is the opportunity to bring price signals closer to customers. Price signals for all customers will not happen all at once; various institutional, technical and cost barriers remain, such as rate structures that do not reflect time variances in the value of electricity and that do not reflect environmental externalities, and the costs of the interface with current wholesale markets, including metering, telemetry, and communication equipment. However, for growing numbers of customers, aggregators such as ESCOs can begin to leverage customer load profiles and the price and quantity data coming from wholesale markets to create value for the customer and the ESCO, and for the market. Combining the economic message of price signals with the environmental message of the value of reducing electricity use is a more powerful incentive for energy efficiency than either message alone.

[2] California and many other states are facing capacity shortages. These are driven in part by uncertainties associated with the transition to restructured markets and the booming economy. Most analysts expect a number of new power plants to come on line in the next few years, which will produce a capacity surplus.

Types of Wholesale Demand Responsiveness[3]

Demand responsiveness provides two types of value to wholesale markets: the value of permanent load reductions and the value of temporary load reductions. Delivery of permanent load reductions looks very much like delivery of permanent load reductions in the retail market, although payment may be structured differently to account for the economics of the wholesale market. Temporary load reductions are similar to wholesale supply bids: the customer (or, more likely, an aggregator) submits a set of price/quantity bids to a wholesale market for a certain time period. Permanent load reductions increase the long-term elasticity of the demand curve, and temporary load reductions increase the short-term elasticity.

Permanent load reductions. Permanent load reductions, such as those resulting from the installation of more-efficient equipment, have economic and environmental value in both the retail and wholesale markets. In the retail market, permanent load reductions reduce kW, kWh, and costs for customers and reduce emissions simultaneously.

People generally think of the economic value of permanent load reductions accruing to the customer.[4] Obviously, if a customer reduces his/her electricity demand, s/he saves a sum equal to the quantity reduced times the price. This is the private benefit. However, permanent load reductions also reduce the price for all other consumers in the unregulated supply market. As shown in Figure 1, when one customer reduces its demand in a particular hour, the entire demand curve shifts down, resulting in a lower price for all customers. This is one measure of the *public* benefit of demand reduction. (Other public benefits include reduced emissions and increased reliability.) (The numbers in Figure 1 are purely illustrative and do not reflect any actual market activity.)

From a public policy perspective, the public benefit of permanent demand reductions could be used to provide incentives to customers and ESCOs to install the technologies and processes necessary to achieve the reductions. This is similar to using public goods charge funds in the retail market to provide incentives for energy efficiency.

It is important to compensate only for actual verified, not calculated, load reductions. The challenge for wholesale demand reductions is to establish methods to measure, confirm, and pay for the energy savings. Various measurement and verification protocols, developed for utility-sponsored demand-side management programs in the retail market (e.g., the International Performance Measurement and Verification Protocol, available at www.ipmvp.org) are also applicable to the wholesale market. ESCOs have a great deal of experience to assist in establishing these measurement protocols.

Stringent measurement and verification techniques are the basis for ensuring that predicted demand reductions actually occur. Measurement and verification are commonly used in the retail delivery of energy efficiency to inform customers of actual energy savings; as a basis for payment for energy efficiency products and services; and to assure regulators

[3] The term demand responsiveness has become much more familiar in the electricity markets in the past year or two, and refers to the willingness of customers (demand) to adjust their level of consumption as the price of the commodity varies.

[4] The discussion in this paragraph is attributable to Rich Ferguson of the Center for Energy Efficiency and Renewable Technologies (Ferguson 1999).

that public funds expended on energy efficiency are delivering appropriate ratepayer benefits. The same protocols can be used in other contexts, such as verifying permanent energysavings in wholesale markets and for the purpose of establishing emissions credits. For example, ESCOs use measurement and verification protocols to design and implement methods for ensuring that the energy efficient equipment the company installs actually produces the expected energy savings and the accompanying reductions of emissions from the electric power plants that would otherwise have generated the saved electricity (Sperberg 2000). Advances in measurement and verification protocols are needed in order to track energy savings by the time periods relevant to the wholesale market, i.e., by portions of hours rather than by the time-of-use periods, days, or months that are adequate for delivery of energy efficiency in retail markets.

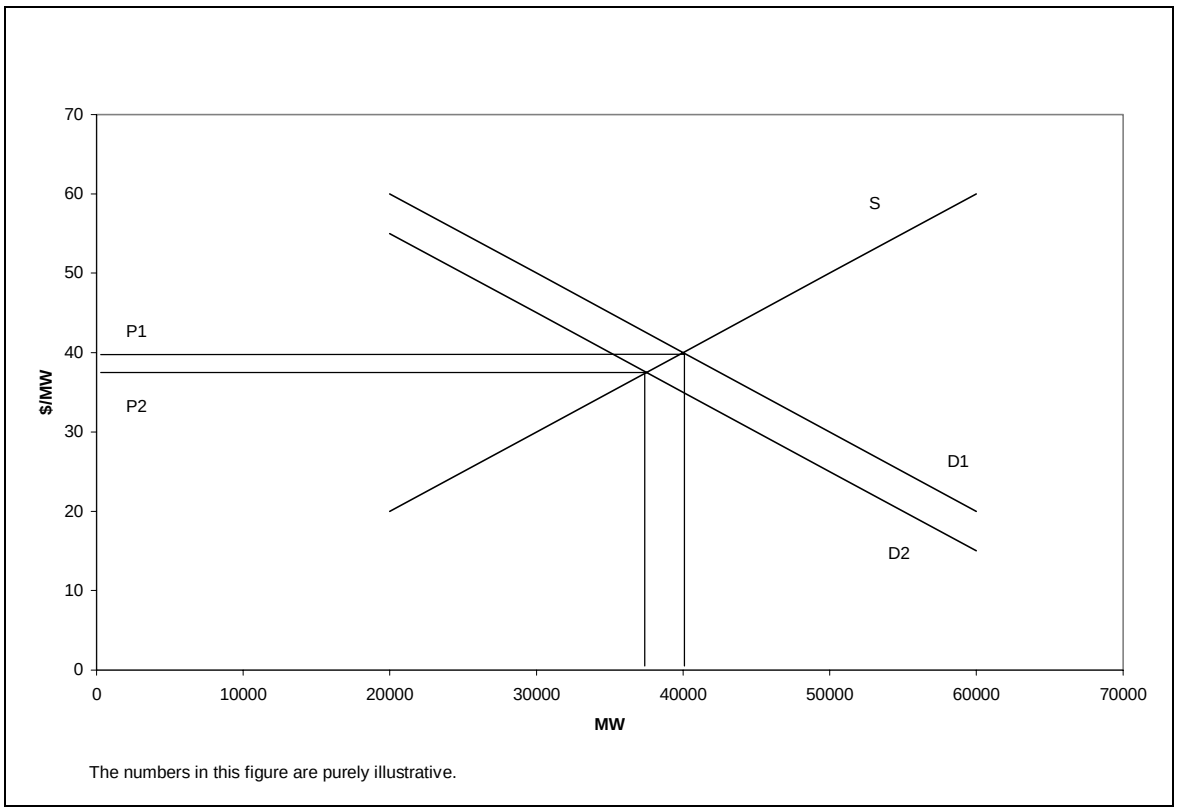

Figure 1: Public Benefit of Price Reduction Due to Demand Responsiveness

Temporary load reductions. In addition to permanent load reductions, wholesale markets can place a high value on temporary load reductions. The problem of a wholesale market is to balance supply and demand. Traditional reliance on supply to fill in any gaps is becoming more expensive as capacity shortages increase. Reducing demand may be quicker, cheaper, and more reliable (due to the diversity benefit of aggregating a number of loads, rather than relying on a single generator) than building new generation.

Temporary load reductions are bid into wholesale markets in the same way that generation is bid into wholesale markets. In California, loads and their aggregators can participate in several wholesale markets. The California Power Exchange (PX) operates day-

ahead, day-of, and block-forward markets. It has also requested permission from the Federal Energy Regulatory Commission (FERC) to operate an ancillary services market. The CAISO operates four ancillary services markets (spinning reserve, nonspinning reserve, replacement reserve, and regulation)[5] and a supplemental energy market (basically a real-time energy market) which provides imbalance energy for uninstructed deviations. It also uses so-called adjustment bids from both generators and loads to solve intrazonal congestion, and has developed an emergency product that would compensate load for voluntary curtailments that occur when Stage 2 system emergencies are called (i.e., when operating reserves drop below Western Systems Coordinating Council requirements).

Generally speaking, load could offer temporary load reductions in the energy markets (PX day-ahead and day-of markets, CAISO supplemental energy market) as well as the capacity markets (primarily the CAISO's nonspinning and replacement reserve markets). In the case of the energy markets, customers (through their aggregators) would indicate a set of price/quantity bids for load reductions (basically a demand curve) for each hour in the day. The Scheduling Coordinator[6] submits the bid to the PX or the CAISO. The PX and CAISO software assembles demand and supply curves for each hour, and the intersection determines the price in that hour. A signal returns from the PX or CAISO to the Scheduling Coordinator indicating the hourly price and the accepted demand reduction level of each customer, given the market price. The aggregator then notifies the customer(s), or exerts automatic control over the load. The customer(s) then is paid only for the load reduction it was willing to provide at that price.

In the case of capacity markets, a capacity auction is added to the process. Customers (through their aggregators) indicate a set of price/quantity bids for load that they would be willing to curtail for each hour in the day (basically a supply curve). The CAISO places these bids in the stack along with generator bids to supply capacity in the same markets. If the load bids win, they receive the auction-determined capacity payment. Then, if the CAISO actually needs that capacity (either nonspinning reserve or replacement reserve), it calls on that load to curtail, just as it calls on generators to supply energy. When that occurs, the customer is paid the market price of energy for each kWh curtailed.

There is some concern that temporary load reductions can harm the environment if the load is not actually reduced, but merely shifted to a different hour. In some regions, peaking generators like natural gas turbines are cleaner than baseload generators like coal plants. The argument is that temporary load reductions merely shift load from relatively clean peak hours to dirtier baseload hours. However, in many regions, peaking units are much less efficient and much less environmentally desirable than, for example, newer gas-fired combined cycle power plants.

The Value of Energy Efficiency to Wholesale Markets

Access to greater amounts of energy efficiency potential through wholesale markets is possible only because the wholesale markets value demand responsiveness. Wholesale

[5] Loads are eligible to participate in the nonspinning and replacement reserve markets.

[6] The California ISO requires that each load and generator submit bids through a Scheduling Coordinator to reduce the number of entities the CAISO must deal with. A Scheduling Coordinator submits a balanced (load and generation) schedule to the CAISO.

markets currently value demand responsiveness because they are capacity-constrained. They may also value demand responsiveness for a number of more theoretical reasons: a more elastic demand curve 1) reduces supplier market power, 2) reduces price volatility, 3) creates a more liquid market, 4) delays the need for new supply, and 5) reduces prices for all customers.

Market Power

In economic theory, one of the conditions for a perfectly competitive market is the existence of a large number of active buyers and sellers who all have full information. In California, both the PX Market Monitoring Committee and the CAISO Market Surveillance Committee have declared that significant supplier market power exists, and have recommended that the PX and CAISO take steps to enhance demand responsiveness. While numerous generators sell into the California markets, there are relatively few buyers. In addition, these buyers tend to bid inelastic demand curves, partly because customers' rates are frozen until the end of the transition period.

The PX Market Monitoring Committee's March 1999 report to FERC states: "We found that during some hours there was considerable potential for generators to exercise market power in the PX market. For example, there were about 100 hours during which offered supply was below predicted end-use demand" (Market Monitoring Committee 1999). The CAISO Market Surveillance Committee similarly found that "significant market power remains in California's wholesale energy market during periods of high total system load" (Market Surveillance Committee 1999). The PX Market Monitoring Committee "showed that demand elasticity could reduce the profit-maximizing bid markup by roughly a factor of four" (Market Monitoring Committee 1999).

Price Volatility

Price volatility, or the existence of price spikes far out of the usual range of prices, can occur when supply becomes so tight that all economic generation units are running, and suppliers believe that customers will pay virtually any price for power. During most months of the year, supply and demand are in close enough balance that prices remain relatively stable. During those moments of capacity constraint, however, prices can go as high as the multi-thousand-dollar-per-MWh prices seen in the Midwest two years ago. Customers do not like price volatility, in part because high electricity prices lead to high operating costs, but also they strongly prefer predictable costs. Thus, clear signals from customers that they are not willing to consume at any price limits the ability of suppliers to set prices sky high.

Liquidity

Liquidity is related to market power. The greater the number of buyers, sellers and trades, the more liquid the market and the greater the opportunity for each actor to participate. The lower the liquidity, the greater the potential for a few actors to exert market power. No matter how many generators there may be, liquidity on the demand side is a desirable discipline on the market.

Energy efficiency is a mature industry in the retail market. It is virtually unknown in the wholesale market. One of the most exciting concepts that has arisen in the past year or two is the notion of adding energy efficiency to the basket of tools of energy traders (Gilbert 1999).

Energy trading is a wholesale function that is almost unknown in retail energy efficiency circles. Energy trading is the buying and selling of the electric commodity at the wholesale level, meaning that trades occur between utilities, marketers, brokers, Electric Service Providers (ESPs), and other market participants, but end-use customers are very rarely involved. They purchase electricity at the retail level from one of the trading entities. Like any commodity traders, energy traders take positions and make and lose significant sums for their employers. Hedging through participation in the forward markets is an important part of the trading function. Options and derivatives are as much a part of trading activity as physical trades. Traders should establish relationships with load aggregators and develop contracting structures that allow the trader to use load reductions, as well as supply, as hedging instruments.[7]

Delay of New Supply

Increasing demand responsiveness sends a more realistic signal to generators. When demand is inelastic and the expectation is that generators will provide every last kWh that customers would like to have, suppliers will build new generation. When demand is elastic, it becomes clear that customers do not want every last kWh, regardless of price, and they will reduce some demand. This delays the need for new generation. If demand responsiveness becomes an accepted and widely-known technique, we may see the U.S. electricity market reaching an equilibrium analogous to that of the world oil market. Over the past two decades, as energy efficiency technologies have become widely known, every price spike in the world oil market has been followed, in short order, by an increase in energy efficiency, especially in the industrial sector. Thus, the spectacular run-ups in world oil prices that were widely predicted in the late 1970s have never occurred, because energy efficiency helped establish a rough market equilibrium.

Price Reductions for All Customers

As noted above, the demand reduction of one customer shifts the entire demand curve down, resulting in lower prices to all customers.

Experience in California To Date

Events and Activities

The original California Public Utilities Commission restructuring decision considered demand responsiveness to be a critical component of creating efficient electricity markets

[7] For a useful discussion of energy trading and hedging, see *Economic Purposes of Commodity Futures Trading*, Commodity Futures Trading Commission, January 1997. www.cftc.gov/opa/brochures/econpurp.html

(California Public Utilities Commission 1995). However, in the effort to implement the PX and CAISO, staff focused virtually exclusively on the supply side. As a result, the PX and CAISO systems, technology, and infrastructure are designed around the characteristics of generators, not load. The key difference between generators and load is the matter of scale: the PX and CAISO systems are designed around generators of 10 MW to 1000+ MW, whereas most loads occur in the kW range. (Though there are several customers with loads in the hundreds of megawatts, they are rare.) The CAISO, in particular, now expects potentially price-responsive demand to comply with technical requirements that are unsuitable for load, as discussed further in a later section.

Events in 1999. Although the CAISO, the PX, and the investor-owned utilities all exhibited some interest in demand responsiveness programs for 1999, none of these programs was particularly successful. The CAISO's Participating Load Agreement (PLA), which is the mechanism for loads to participate in some of the CAISO's ancillary services markets and its supplemental energy market, was implemented by only one customer in summer 1999. Neither PG&E nor SCE implemented its demand responsiveness program after the California Public Utilities Commission (CPUC) approved the programs but with significant modifications. At the PX, it is technically feasible for a customer's scheduling coordinator to submit a demand curve (i.e., a set of price/quantity bids), as opposed to a hard-and-fast demand number. However, this practice is rare.

In summer 1999, the CAISO created the Participating Load Agreement, which is now part of its FERC-approved tariff. The PLA defines the terms under which load can participate in these markets. Unfortunately, the PLA's requirements are such that almost no customers in California have been willing and able to sign a PLA and operate in the ancillary services market. The technical requirements, such as the 4-second scan rate, are so stringent, onerous, and inapplicable to load that almost none have come forward.

In December 1999, the CAISO issued Draft Technical Standards for load participation in its markets. This document attempted (but failed) to create standards that met the perceived requirements of the CAISO and also encouraged load participation. Market participants' comments on the Draft Technical Standards pointed out that the standards were still so stringent that virtually no load could meet them.

Events in 2000. As noted, the PX and CAISO were originally designed with generation in mind. It wasn't until the CAISO saw its summer 2000 peak demand forecast that it began a more serious effort to solicit load participation in its markets.[8]

A May 10, 2000 CAISO news release cited an expected normal-weather peak load of 46,250 MW, which would be met by 46,350 MW of in-area generation and net imports

[8] A secondary impetus was the concerns over market power expressed by both the PX Market Monitoring Committee and the CAISO Market Surveillance Committee. At the CAISO, the issue of demand responsiveness is directly relevant to the question of price caps, which in turn is of major interest to suppliers and customers. The CAISO currently has a price cap of $750/MWh. The price cap was revisited in March 2000; if certain conditions had not been met, the price cap would have been reduced to $500/MW. One of those conditions was evidence of increased demand responsiveness.

(CAISO 2000). After a May 22, 2000 Stage Two emergency occurred, CAISO staff concluded that 48,000 MW would be the new peak load forecast (Savage 2000).

The CAISO now has three summer 2000 load programs and, for 2001, has announced load participation in a reliability program that previously used only generation. The CPUC has approved demand responsiveness programs for Pacific Gas and Electric (PG&E) and Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E) has proposed two demand responsiveness programs upon which the CPUC has not yet acted.

The first CAISO demand responsiveness program is an Invitation to Bid into the Ancillary Services and Supplemental Energy Markets For Summer 2000. The CAISO issued the Invitation on February 29, 2000 and signed the first PLAs in mid May.

In this program, participants consist of aggregated loads that have the ability to bid into the CAISO's nonspinning reserve and replacement reserve (ancillary services) markets and supplemental energy market on a daily basis. For this program the CAISO is requiring loads to adhere to its Technical Principles, which require stringent metering, telemetry, and communications infrastructure. (See section on information and technologies.) In its Invitation, the CAISO limited participation to 400 MW for the nonspinning reserve market, 400 MW for the replacement reserve market, and 1000 MW for the supplemental energy market. In reality, the CAISO received five or six bids for 400-500 MW in all three markets combined.

The second CAISO summer 2000 demand responsiveness program is its Demand Relief Program (DRP). The DRP targets nonresidential customers whom the CAISO will compensate for shedding load when the CAISO calls them, generally after it declares a Stage One emergency (i.e., when reserves drop below 7%). The DRP is a reliability product, not a market product. The DRP is in some respects similar to the utilities' interruptible load programs, and is designed in part to help transition from those programs to the future. California statute requires the interruptible credit to expire in March 2002.

Participating customers responded to a Request for Bids issued in the spring and signed Demand Relief Agreements (which are different from Participating Load Agreements) in time for the June 15 start date of the program. Compensation includes a capacity payment for the period June 15 - October 15, 2000, and an energy payment for the hours when the CAISO issues a dispatch notice – presumably after it calls Stage One emergencies. The capacity payment is the amount the customer bid to the CAISO and is not a market price, e.g., it is not based on an auction. The energy payment is based on the real-time energy price. Customers in this program must have interval meters capable of recording usage on an hourly basis, but do not need to comply with the CAISO's telemetry and communication requirements.

In the DRP, the CAISO had hoped to receive bids for up to 1000 MW of load that would be willing to curtail during and after a Stage One emergency. In actuality, the CAISO received 67 bids for a total of 200+ MW and conditionally accepted bids for 180 MW. The smallest bid was for 1 MW (the minimum allowed) and the largest was for 50 MW. In addition, the prices that participating loads bid were somewhat higher than the CAISO anticipated. The average bid price was $37,739/MW-month, with a low of $500/MW-month and a high of $120,000/MW-month. The high bid of $120,000/MW-month corresponds to bidding the price cap of $750 for the 160 hours of the month when load must commit to

being capable of curtailing.[9] The total cost to the CAISO of this program's capacity payments is $26 million. (Upon receiving bids, the CAISO suspended the bid process for about a week because of concern over the price and quantity bids, but ultimately recommended that the CAISO Board of Governors approve the program, which they did.) The program cost will be recovered from grid customers.

The third CAISO summer 2000 demand responsiveness program is Power Watch 2000, which is a public awareness program which will give consumers 24-hour notice of expected high-demand days. Customers are asked to voluntarily reduce load. The Power Watch 2000 news alerts will give tips on how to conserve energy on those days.

As noted above, SCE and PG&E also have summer 2000 demand responsiveness programs for their bundled customers. The CPUC recently approved a Voluntary Power Reduction Credit for SCE and Schedule E-BID for PG&E. In these programs, the utility notifies participating customers when it expects the PX price to be greater than $250/MWh. Customers have the choice to curtail or not. If they curtail, they receive an incentive payment based on the PX day-ahead market clearing price. Customers that commit to curtail on a particular day and then do not curtail pay no penalty but would be removed from the program after two or three failures to curtail. The costs of the program include administrative costs and the cost of the incentive payment; these costs will be recovered from all bundled ratepayers.

Several energy service providers and other parties protested the utility programs on the basis that the programs are anticompetitive and prejudge the role of the utility in a restructured market. The CPUC acknowledged these concerns and stated that these programs set no precedents for programs beyond summer 2000. (The CAISO has also made it clear that its programs are for summer 2000 only and will be re-evaluated for future years.)

Customers with interval meters and demand greater than 500 kW and 50 customers with lower demand are eligible. The programs are limited to 500 customers and/or 500 MW of load per utility. The CPUC added several requirements for studying the costs and impact of the program on the price elasticity of demand, on other ratepayers, and on the wholesale market price.

For 2001, the CAISO has already issued its Request for Bids for the 2001 Local Area Reliability Service (LARS) process. Bids were due July 1. Through the LARS process, the CAISO would execute a Reliability Must-Run (RMR) contract with the load or its aggregator. There is a 10 MW minimum. RMR contracts involve substantial capacity payments and allow the resource under contract to participate in various energy markets as well.

Observations

Scale. As noted above, a major difficulty with CAISO demand responsiveness programs has been that the CAISO believes that load should look and act just like generation, particularly in the types of operational and settlement data that are provided. The CAISO has struggled to comprehend the fundamental scale distinction between load and generators. Specifically, the CAISO originally required loads to have telemetry installed capable of providing 4-second

[9] A MW-month is based on load being available to the ISO for eight hours per day, five days per week, or 160 hours/month, although the ISO can call the load for a maximum of 30 hours per month.

scan rates for the CAISO's energy management system, and required similarly stringent metering and communications technology.

The existing technical requirements for load participation were modified from the requirements for generators. The fundamental difference between load and generators, from an operational point of view, is scale. Generators are typically 50-850 MW. By contrast, a very large load is 10 MW. More typically, aggregated loads that could contribute to demand responsiveness in the California markets are measured in kW, with blocks of >1 MW being available for bidding. From the point of view of balancing supply and demand in California's 48,000 MW system, it is important to know the status of a 50+ MW generator, and the 4-second scan rates, etc., are appropriate for ensuring reliable operation. However, it is entirely unnecessary to monitor each load in a 1-MW aggregation with the same precision, since the presence or absence of any one load in the aggregation will not throw the system significantly out of balance. In addition, if one load in the aggregation turns out to be unreliable, it does not affect the availability of the rest of the aggregation. Thus, aggregated loads are actually more reliable than generation, due to their diversity.

Institutional issues. Although the CAISO, PX and investor-owned utilities are making efforts to improve demand responsiveness, a number of institutional barriers remain. These include the complexity of technical and contractual requirements; price caps in the ancillary services markets; the CAISO's 10-minute market; and the eligibility of customers on interruptible tariffs.

An early barrier to customer and ESCO participation in demand responsiveness programs is the complexity of the technical and contractual requirements to participate. For example, the CAISO's PLA requires that an aggregator provide the meter number, location, and other information for each meter in its aggregation. Adding or deleting a meter requires 60 days notice, and is not effective until the next CAISO update of its master data file. While this requirement is no doubt sensible from the CAISO's perspective, it appears unnecessary and burdensome from an aggregator perspective, since it drastically reduces the aggregator's flexibility in choosing which loads to include at any given time.

The CAISO also requires the use of scheduling coordinators, which are entities that perform many of the bidding, scheduling, metering, billing, and settlements functions in California. Not only must an aggregator pay for a scheduling coordinator's services, but the scheduling coordinator is an extra step between the load and the market. In addition, if the aggregator wants to bid into the ancillary services markets, only a scheduling coordinator with an ancillary services certificate will do.

Technical requirements for participation in the ancillary services markets address the CAISO's wish to "see" load at every moment in its energy management system, as if it were a generator. As discussed, these requirements are inappropriate for loads due to the difference in scale between loads and generators. The technical requirements add tremendous complexity and cost in terms of equipment purchase and installation and data manipulation and storage. These costs add significantly to an aggregator's transaction costs, potentially to the extent of eliminating any benefit of participating in the programs.

The Demand Relief Program offers the substantial benefit of not requiring telemetry and communications equipment. For this reason alone, many customers and aggregators may prefer this program, which is a reliability program designed to address emergency situations,

to participation in the ancillary services markets. This is unfortunate to the extent that it reduces market liquidity by shifting market load into a reliability product.

As of this writing, the CAISO price cap is $750/MWh. The cap was imposed to protect market participants while market design details are worked out. However, the CAISO cap (which also acts as an effective cap in the PX energy markets[10]) is a strong inhibitor of aggregator participation in wholesale markets. Aggregators face significant costs to gather customers, install metering, telemetry, and communication equipment, and to set up the operational and financial structures necessary for interacting with the PX and CAISO. They also know that, for many hours of the year, supply and demand are sufficiently balanced to discipline market prices, and that there are relatively few hours each year when prices can rise sufficiently to cover the aggregators' costs. They must also share any profits from these hours with customers, who, depending on the curtailment technology installed, demand substantial compensation for the inconvenience of diminishing, interrupting, or delaying their processes. A cap of $750 unnecessarily dampens customer and aggregator interest in providing demand responsiveness. The price cap should be reconsidered for providers of demand responsiveness.

Another impediment to demand responsiveness is the CAISO's 10-minute markets. In an effort to solve other market problems, in August 2000, the CAISO will begin settling its markets in 10-minute increments rather than one-hour increments. That is, it will charge and pay loads and generators based on their activities and the price within each 10-minute period. This feature drastically reduces the amount of load that can be demand responsive, since most load is not capable of ramping up and down every 10 minutes in response to CAISO commands. One-hour increments are much more suitable for the characteristics of load.

Finally, another major issue for all programs has been whether customers that are currently on utility interruptible rate schedules are eligible. In theory, these customers' load is already counted towards the state's operating reserve, and is already receiving payment for willingness to curtail in the form of the nonfirm credit off their otherwise applicable tariff. On that basis, they should not be eligible. On the other hand, these are the customers in the best position to curtail when necessary. At this time, interruptible customers are eligible for the CAISO's DRP, pending CPUC approval; ineligible for the ancillary services and supplemental energy markets; and eligible for the SCE and PG&E programs.

Information and Technologies Used to Deliver Retail and Wholesale Energy Efficiency

The main differences in the information and technologies used to deliver retail and wholesale energy efficiency are in the frequency and precision of the measurement of energy savings, the communications requirements, and the frequency of data collection for operational purposes.

Retail providers of energy efficiency are most concerned with having the data that justifies the installation of a project, the data that enables them to maintain the equipment and

[10] The ISO price cap of $750/MWh is not always an implicit price cap in the PX markets. On May 22, 2000, the PX day-of /hour-ahead market price for hour 17 was $863/MWh (O'Donnell 2000). May 22 also saw an ISO-declared Stage Two emergency.

ensure ongoing energy savings, and the data that verifies actual energy savings. However, wholesale markets require information in the form of operational data that are used to balance supply and demand in real time.

Frequency and Precision of Verified Load Reduction

For retail markets, providers use customer billing information, often in aggregate, to identify potential customers, to help establish baseline energy use, to assess which energy efficiency technologies are most appropriate for a particular customer, and to charge customers for energy efficiency products and services. Billing information generally includes monthly usage, monthly total bill, and sometimes includes time of use data and demand data. Although time-of-use and interval meters are very useful, many retail energy efficiency customers have only the simplest of meters installed. One strategy for delivery energy efficiency is for an ESCO to install submeters within a customer facility in order to measure and verify energy savings.

In the retail market, little information flows between ESCO and customer on the one hand and the utility, the PX, and the CAISO on the other hand, because the transactions are between the ESCO and the customer and occur on the customer side of the meter. The only concern the utility (or energy service provider, for a direct access customer) has is that it be able to forecast load reasonably accurately on an hourly basis, because the CAISO charges for deviations from forecasted load. Permanent load reductions on the retail side, while potentially helping to alleviate transmission and distribution bottlenecks, usually do not affect the real time operation of the transmission or distribution systems in ways that require sophisticated metering and/or communications systems.

For temporary load reductions in the wholesale markets, entirely different information is required. From the ESCO and customer point of view, the relevant data include predicted and actual hourly capacity and energy prices, which allow the ESCO and customer to make informed bids. From the CAISO's point of view, financial considerations are secondary to the CAISO's obligation to operate the electric system and to keep supply and demand in balance. Thus, the CAISO desires data on real-time operations; it wants to know, ideally within 4 seconds, exact load levels. The CAISO would also prefer to receive data such as kVAR from the load on a continuous basis.

Since the wholesale markets settle on an hourly basis, and will begin to settle on a 10-minute basis in August 2000, the CAISO also wants load data in hourly and preferably 10-minute increments. These CAISO requirements are far more rigorous than those required by the contractual arrangements of retail markets, which often require only billing data on a monthly basis.

Communication Requirements

Delivery of energy efficiency at the retail level requires communication systems in which individual meters or submeters transmit data to a central energy management system (EMS), and receive control signals from that EMS. The EMS can be located either at the facility or at some remote control center operated either by the facility owners or a third-party contractor. The communications systems are frequently based on direct links: dedicated

phone lines, cell phones, satellites or radio frequency technology. Internet communication is now often substituted for direct links because of its greater speed and economy. The main purpose of these communication links is to manage facility operations to maximize the value of the facility's energy use.

For temporary load reductions in wholesale markets, more sophisticated communications technology is currently required because of the need to balance supply and demand instantaneously and continuously. Data must be transmitted much more frequently, and over highly reliable and extremely secure networks. Redundant communications links can be important. The California ISO is setting up secure Internet connections for its communication links. In wholesale markets, dedicated phone lines are being supplanted by value-added networks (VANs). The question is whether the current communications requirements for the wholesale markets should be extended to load reductions, or whether a less rigorous communications protocol will suffice.

Conclusions

Once restructuring electricity markets establish the technical, economic, and institutional mechanisms for encouraging demand responsiveness, hundreds of megawatts of both permanent and temporary demand reductions will become available to the wholesale markets. This will provide enormous benefits to the wholesale markets (both operators and customers); provide another avenue for accessing additional potential for energy efficiency, with its attendant environmental benefits; and benefit the ESCO industry. Synergies between retail and wholesale benefits of energy efficiency will push more potential projects into cost effectiveness.

References

Carter, S. 2000. *Investments in the Public Interest: California's Public Benefit Programs Under Assembly Bill 1890.* San Francisco: NRDC.

California Independent System Operator. May 10, 2000. *News Release: Surge in Electricity Consumption Squeezes Surplus Power: California Initiates Power Watch 2000 Campaign.*

California Public Utilities Commission. December 20, 1995. Decision 95-12-063.

Ferguson, R. 1999. "The Value of Energy Efficiency in Competitive Power Markets." *Clean Power Journal* Summer Edition, 1999.

Gilbert, Joel. 1999. Preworkshop Training, Capitalizing on the Peak Load Management Opportunity, West Palm Beach, Fla., October 13, 1999.

Hirst, E. 1991. *Possible Effects of Electric-Utility DSM Programs, 1990-2010*. ORNL/CON-312. Oak Ridge, Tenn: Oak Ridge National Laboratory.

Kirby, B. and E. Hirst. 1999. Load as a Resource in Providing Ancillary Services. Paper presented at the American Power Conference, Chicago, Illinois, April 1999.

Market Monitoring Committee of the California PX. March 9, 1999. *Second Report on Market Issues in the California Power Exchange Energy Markets.*

Market Surveillance Committee of California Independent System Operator. 1999. *Report on Redesign of California Real-time Energy and Ancillary Services Markets.*

O'Donnell, A. May 26, 2000. "Market Emergency Leads to New Price Records at CalPX." *California Energy Markets* No. 568, p. 6.

Savage, J.A. May 26, 2000. "Some Customers' Power Cut While Cal-ISO Deals With Heat Wave, Outages." *California Energy Markets* No. 568, p. 9.

Sperberg, R. March 2000. Personal Communication.

Acknowledgements

The authors would like to acknowledge the contributions of Richard T. Sperberg, Terry E. Singer, Dian M. Grueneich and Leslie Kramer to this paper. This paper has also benefited from the attention of careful reviewers.