Energy 35 (2010) 1536–1543

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Demand participation in the restructured Electric Reliability Council of Texas market

Jay W. Zarnikau [a,b,*]

[a] Frontier Associates, 1515 S. Capital of Texas Hwy., Ste. 110, Austin, TX 78746, USA
[b] LBJ School of Public Affairs and College of Natural Sciences, Division of Statistics, University of Texas at Austin, 4131 Spicewood Springs Road, Suite O-3, TX 78713, USA

ARTICLE INFO

Article history:
Received 23 January 2009
Received in revised form
28 February 2009
Accepted 5 March 2009
Available online 8 May 2009

Keywords:
Demand response
Restructured electricity markets
ERCOT

ABSTRACT

Does an electricity market which has been restructured to foster competition provide greater opportunities for demand response than a traditional regulated utility industry? The experiences of the restructured Electric Reliability Council of Texas (ERCOT) market over the past eight years provide some hope that it is possible to design a competitive market which will properly value and accommodate demand response. While the overall level of demand response in ERCOT is below the levels enjoyed prior to restructuring, there have nonetheless been some promising advances, including the integration of demand-side resources into competitive markets for ancillary services. ERCOT's experiences demonstrate that the degree of demand participation in a restructured market is highly sensitive to the market design. But even in a market which has been deregulated to a large degree, regulatory intervention and special demand-side programs may be needed in order to bolster demand response.

1. Introduction

Progress in facilitating demand response in Electric Reliability Council of Texas (ERCOT) market may be of particular interest for a variety of reasons. The restructuring and introduction of competition into ERCOT have been relatively successful [1,2] and a responsive demand side is often viewed as a prerequisite to the establishment of a successful competitive electricity market [3]. ERCOT is a highly unbundled market with considerable separation among generation, transmission, and retail functions. The very high levels of participation by industrial energy consumers within this market on interruptible and real-time pricing tariffs prior to restructuring leaves no question that a very large number of loads are capable of responding to either price signals or curtailment requests. Thus ERCOT's experiences can provide valuable insights into the challenges and opportunities associated with fostering demand response in a highly restructured market.

About 85% of the electricity needs in the largest electricity-consuming state in the U.S. are satisfied through the ERCOT market. ERCOT contains over 550 generating units, with a combined capacity of roughly 90 GW. This market represents about 10% of the total electricity sales in the U.S. Natural gas is used for over 44% of the electricity generation involved in market transactions (and probably over 50% of the total when on-site generation is considered) and constitutes well over 70% of the generating capacity. More detailed descriptions of the ERCOT market may be found in Adib and Zarnikau [4]. Because ERCOT is an 'intra-state' market with limited interconnection to other markets in the U.S., there is very little federal regulatory jurisdiction over ERCOT. This unique jurisdictional status has permitted the Texas market to adopt policies and market features which diverge from other markets.

ERCOT has undergone gradual restructuring since the mid-1990s to foster competition in the wholesale and retail segments of the industry and to relax regulatory oversight. Senate Bill 373, enacted in 1995, required the Public Utility Commission of Texas (PUCT) to establish rules to foster wholesale competition and create an Independent System Operator (ISO) to ensure non-discriminatory transmission access, leading to ERCOT's establishment as the first operating ISO in the U.S. in the summer of 1997. Sweeping reforms resulted from 1999's Senate Bill 7 (SB 7) which allowed customers of the investor-owned utilities within ERCOT to choose among various retail electric providers (REPs) – load-serving entities which provide a retail supply of electricity within the areas of ERCOT opened to retail competition – for a retail supply of electricity beginning on January 1, 2002. Over 100 REPs presently provide a retail supply of electricity to consumers, including REPs established to serve individual industrial loads. Roughly 15 REPs actively serve the residential customer segment. SB 7 also provided the ISO with much greater centralized control over the wholesale

* LBJ School of Public Affairs and College of Natural Sciences, Division of Statistics, University of Texas at Austin, Frontier Associates 1515 S. Capital of Texas Hwy, Austin, TX 78746, USA. Tel.: +1 512 372 8778; fax: +1 512 372 8932.
E-mail address: jayz@frontierassoc.com

0360-5442/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2009.03.018

market and led to the establishment of formal markets for ancillary services and balancing energy [4].

Numerous industrial facilities involved in chemical production, petrochemicals, refinery operation, air separation, pulp and paper manufacturing, and steel production reside in Texas. Overall, industrial energy consumers account for over one-quarter of the electricity purchases in the market. Many of these facilities can withstand short interruptions in their electricity supply with modest economic loss. Traditionally, these facilities were served through interruptible tariffs, which provided an electrical supply to the facility at a lower level of reliability in return for a discounted price. Consequently, there is little doubt that a significant share of the demand side of the ERCOT market has the capability to reduce or curtail electricity purchases in response to either an instruction from the ISO or in response to a price signal [5].

Prior to the introduction of retail competition in January 2002, ERCOT relied upon roughly 3500 MW of interruptible load, group load curtailment programs, residential direct load control, and other load management programs to maintain reliability [6]. Innovative pricing programs had also proven successful. As the ERCOT market was redesigned between 1999 and 2001 to foster competition, much of this large demand-side resource was lost, at least temporarily. Restructuring required the termination of all tariffs in the areas of ERCOT opened to retail competition, including the tariffs offering a discounted price to interruptible loads and the tariffs used to provide consumers with real-time pricing options. Thus, on January 1, 2002, the market lost a planning reserve resource of nearly 3000 MW. Also lost was under-frequency response from large industrial loads on instantaneous interruptible tariffs which was used to offset spinning reserves requirements under the previous utility structure.[1] Further, the restructuring plan required utilities to divest their "competitive energy services," leading to the termination of residential load management programs. Confusion ensued over who would assume responsibility for overall resource adequacy.

In hopes of preserving some of the demand response capability which had been developed in the 1980s and 1990s, the PUCT ordered ERCOT to "develop new measures and refine existing measures to enable load resources a greater opportunity to participate in the ERCOT market" [7]. This directive was re-affirmed by the PUCT in subsequent orders [8] and preserving this resource became increasingly important as the PUCT gradually increased offer caps in the ERCOT wholesale market and adopted an "energy-only" resource adequacy mechanism. The measures adopted by ERCOT and their degree of success are reviewed in this paper.

2. Overview of demand response opportunities in ERCOT

While the PUCT provides broad policy guidance, the development of ERCOT's detailed rules and procedures rely upon a stakeholder process, as codified in the ERCOT Protocols [9]. The appropriate role of demand response in the new market was (and still is) hotly debated. But there was some agreement among stakeholders with the general principle that demand-side resources should be permitted to compete "head-to-head" with generation resources in ERCOT-operated markets, provided the demand-side resource could provide a service similar to the services provided by a supply-side resource and reliance upon demand-side resources in lieu of a supply-side alternative would not compromise reliability. Supply and demand-side resources are not equivalent. But demand-side resources would have an opportunity to provide a service, provided the demand-side resource could "act as" a supply-side resource.

In addition to the opportunity to formally participate in markets for energy and ancillary services administered by ERCOT, the market settlement process was designed such that price-responsive loads could potentially be rewarded. Also, through the balancing up load (BUL) program, price-responsive loads agreeing to formally offer their price response to ERCOT by submitting offers to provide an offset to balancing energy would qualify for an additional economic incentive. In 2007, an emergency curtailment program was added to the menu of demand-side participation opportunities.

Table 1 categorizes demand-side participation which directly involve ERCOT's markets or settlement process. In addition to those listed here, a number of demand response programs are offered on a bilateral basis between load-serving entities and consumers, and may not be visible to the market.[2]

Each of these opportunities for encouraging demand response will now be reviewed, beginning with Loads Acting as Resources (LaaRs) in ERCOT's markets for ancillary services (Table 2).

3. Using interruptible loads as an ancillary service

The first challenge faced was defining an appropriate role in new competitive markets for ancillary services for the large industrial energy consumers with loads that were formerly served under interruptible tariffs. It was decided that LaaRs could directly compete against generation resources to provide various ancillary services, including Responsive Reserve Service (provided by interruptible loads with under-frequency relays), Non-Spinning Reserve Service (which can be interrupted by the ISO with 30 min of notice), and Regulation (requiring a response similar to a power plant's governor). There are some ancillary services (e.g., black start) which a demand-side resource obviously cannot provide.

LaaRs selected to provide ancillary services through ERCOT's formal day-ahead markets receive the market-clearing price for capacity. Alternatively, LaaRs may be self-arranged by a load-serving entity, in which case they would receive a negotiated price.

The restructured ERCOT market has probably been as successful as any market in the integration of interruptible loads into markets for ancillary services. As of the end of 2008, 144 LaaRs (working with 10 scheduling entities) are qualified to provide ancillary services for a total capacity of 1984 MW. The amount of load qualified to provide Responsive Reserves is well in excess of the constraint on LaaR participation in this market (discussed below). LaaRs have been instrumental in preserving reliability, and typically three interruptions occur per year in response to under-frequency events. While there has been a good mix of LaaRs by size, it is noteworthy that about one-half of the total quantity of LaaRs is provided by five very large industrial loads (Table 3).

One LaaR (associated with a facility owned by Oxy, which also has generation) began providing regulation in late 2006 under a "Controllable Load" pilot program. Beginning in March 2007, LaaRs were also permitted to provide replacement capacity. Demand-side resources have been proposed as substitutes for reliability-must-run generating units, but no proposals have been

[1] A small amount of interruptible load remained in the service areas of municipal utility systems and rural electric co-operatives which were not affected by retail competition.

[2] In addition to the programs and activities mentioned here, there are a number of small demand response programs in ERCOT which have nothing to do with the competitive market. These include Austin Energy's direct load control program, interruptible tariffs and programs offered by "non-opt-in entities" which include most of the municipal and rural electric cooperative systems in the ERCOT region, and the Load Management Standard Offer Programs which are regulated energy efficiency programs offered by transmission and distribution utilities including Oncor Electric Delivery and AEP-Texas Central.

Table 1
Types of demand response in ERCOT[a].

Resource type	Service that can be provided	Requirements	Approximate quantity in 2008 (MW)	Approximate cost in 2008
Voluntary Load Response (VLR)	Curtailment or reduction in response to market price or other factors	• Metering and/or curtailment technology defined in REP contract	600	Not applicable
Qualified Balancing Up Load (BUL)	BUL (participation in the balancing energy market)	• IDR meter • ERCOT qualification, which involves registration.	0	Not applicable
Load Acting as a Resource (LaaR)	Various ERCOT ancillary services (A/S)	• IDR meter • Telemetry • ERCOT qualification, which involves registration, telemetry, and testing	2000 (registered) 1200–1300 (typical level of participation)	$15 per MWh
Emergency Interruptible Load Service (EILS)	Curtailment in response to ERCOT dispatch	• IDR meter or non-IDR aggregations • ERCOT qualification, which involves registration, review of baseline calculations, and testing.	275	$10 per MW

[a] Source: ERCOT Demand Side Working Group (2006). Load Participation in the ERCOT Market.

accepted to date. The use of demand-side resources to manage transmission congestion has been very limited.

The amount of LaaR that can be used to provide Responsive Reserves has been subject to caps. There is a concern that if too large a share of responsive reserve requirements were provided by LaaRs, then there might not be adequate generation resources providing responsive reserves. Generating units with governors are better able to stabilize frequency in response to small deviations in frequency than LaaRs with their *off-or-on*, *all-or-nothing* response. Also, machines with physical *mass* are needed to maintain the stability of the network.[3] Further, there is a concern over the possibility of "over-shoot," where too much interruptible load might trip-off at the same time and raise frequency to an unacceptably high level, although this concern can be addressed by setting under-frequency relays to a mix of trigger points. Initially, this limit was 25% of ERCOT's requirements for this ancillary service (i.e., 575 MWh, given ERCOT's normal requirement of 2300 MW). Later this cap was raised to 1150 MW (which was one-half of ERCOT's normal need for responsive reserves until January 2008, when overall Responsive Reserve requirements began to vary between 2300 MW and 2800 MW), as concerns surrounding over-shoot abated [10]. In addition, strict qualification criteria were introduced to preclude energy consumers whose load level could not be accurately predicted on a day-ahead basis from providing responsive reserves. Within a couple of years after LaaRs were permitted to participate in the wholesale market, the 1150 MW cap was reached.

The cap, coupled with restrictive qualification requirements, has led to situations where interruptible loads have been willing and able to interrupt in order to balance supply and demand during reliability problems (e.g., an emergency event on April 17, 2006), but could not be deployed since they were not selected by ERCOT to provide an ancillary service at that time or could not meet ERCOT's stricter qualification criteria.

The excess supply of LaaRs relative to the cap has led to problems. As competition among LaaRs intensified for their limited share of the market for responsive reserves, many LaaRs began offering their interruption capability at increasingly negative prices in hopes of securing a place among selected resources within the bid stack, and in anticipation that a higher-price generation resource (which is unlikely to bid at a price below its positive start-up and fuel cost) would set the market-clearing price which all selected resources receive. This bidding behavior is depicted in Fig. 1. There could be severe consequences if an offer price as low as −$19,155 per MW actually set the market-clearing price. In such a case, all of the selected resources would then have to *pay* the market that price. This credit risk led to the imposition of a temporary floor price to prohibit negative bids. It also led to discussions over the merits of creating separate markets for LaaRs and generators providing responsive reserves.

Overall, a reasonable balance has been achieved. LaaRs typically provide 1150 MW of Responsive Reserves and this competition undoubtedly lowers the overall cost of providing that operating reserve relative to the costs which would result if only generators were eligible to offer the service. LaaR participation in other markets for operating reserves has been limited. But, opportunities for demand-side participation in these other ancillary service markets are at least available to qualifying loads.

While some complain that requiring interruptible loads to "act as" a supply-side resource demotes demand-side resources to second-class status, there certainly are many differences among the two types of resources with respect to the nature of their response to deployment requests, their bidding behavior, and the impact of their deployment on system operations. In some circumstances, demand-side resources can provide a superior service. For example, the response of an under-frequency relay may be faster than the ramp-up of a power plant in addressing a drop in frequency.

The new Nodal Protocols, which will dictate the market rules when ERCOT eventually adopts a nodal locational marginal cost wholesale market structure, replace the term LaaR with demand resources, to suggest an "equivalent" value of demand-side resources compared to supply-side alternatives [11]. Yet physical differences will remain. The degree to which demand and supply-side resources are truly equivalent is often debated. And some programs specifically for demand-side resources have been initiated, as noted below.

4. Load response to price signals and programs offered by REPs

A second challenge was to ensure that industrial energy consumers who could readily respond to price signals and that were formerly served under certain interruptible or real-time pricing tariffs would continue to be provided with opportunities to curtail or shift power purchases in response to price signals. Such responses would assist the overall system in matching supply and demand, reduce the need to additional generating capacity, and provide price-responsive consumers with opportunities to reduce their energy costs. ERCOT's market structure was designed to provide incentives for large energy consumers to reduce power purchases during peak or high-price periods.

The design of transmission charges is based upon the consumer's contribution to demand during four peak times in summer

[3] The inertia of a power system or its ability to maintain a consistent frequency is directly related to the size (mass) of the generation units and the number of generators on line.

Table 2
LaaR deployment summary – 2005 to present[a].

Date	Time	Duration (min)	Account of response (MW)	Type of deployment
2005				
2/13/2005	18:35	60	363	Under-frequency Spike and Operator frequency restoration by verbal dispatch
7/1/2005	15:13	84	178	Mitigation plan related to local congestion/transmission security
2006				
4/17/2006	15:34	175	1357	Emergency curtailment Plan Step 2, Systemwide manual deployment
10/3/2006	17:37	137	1291	Systemwide manual deployment related to frequency restoration
12/22/2006	2:54	64	1215	Under-frequency Spike < 59.7 but of uncertain duration followed by ERCOT Operator Verbal dispatch
2007				
7/2/2007	19:38	25	1184	Systemwide manual deployment related to frequency restoration
9/3/2007	7:57	92	1206	Systemwide manual deployment related to frequency restoration
12/12/2007	1:56	19	1104	Systemwide manual deployment related to frequency restoration
2008				
2/26/2008	18:49	79	1211	Emergency curtailment Plan Step 2, Systemwide manual deployment
3/16/2008	11:37	4	632	Under-frequency event frequency < 59.7 Hz but less than 20 cycles
8/11/2008	17:14	41	1179	Systemwide manual deployment related to frequency restoration

[a] Source: Steve Krein of the ERCOT Staff (2008). "Load Participation Update." Demand Side Working Group Meeting, December 12.

months (4 CPs), thus providing large consumers with an incentive to reduce their power purchases during the summer peaks. Often, transmission charges are treated as "pass-through" costs in the contracts offered by REPs. Consequently, larger energy consumers may see direct benefits by reducing their consumption during the four summer peaks, which are used to allocate transmission costs to consumers and load-serving entities.

The ability to purchase balancing energy (through a REP) and the design of the wholesale market settlement system reward qualified scheduling entities (QSEs) who can reduce generation needs during high-price periods. Consumers are free to deviate from scheduled load levels (in response to price changes, for example) with minimal penalties. "Voluntary load response" refers to a customer's deviation from its scheduled or anticipated load level in response to price signals (e.g., balancing energy prices or peak demand periods used to assign transmission costs) in situations where the customer has not formally offered its response to the market as a "resource." If the actual load level of a QSE turns out to be lower than its scheduled load level during a given 15-min interval while its actual generation is equal to its scheduled generation, then the QSE is entitled to a payment or credit based on the energy imbalance multiplied by the balancing energy market price. This may provide energy consumers with an incentive to respond to wholesale market prices, provided their REP agrees to settle the consumer separately from other loads served through the REP. This separate settlement normally requires the metering of a load's consumption at 15-min intervals. At market-open in 2002, interval data recorders (IDRs) were required on energy consumers with a billing demand over 1 MW. The IDR threshold was later reduced to 700 kW.

Some industrial energy consumers rely on balancing energy (essentially, spot market power) to meet some or all of their electricity needs. These consumers actively monitor 15-min balancing energy prices, and reduce electricity purchases when prices exceed threshold levels. The degree to which this practice is permitted has been subject to changing policies since the introduction of customer choice. During the first years of the restructured market, there was a "balanced schedule requirement" (although some load-serving entities ignored it). Later, a "relaxed balanced schedule requirement" was introduced, in part to encourage REPs and large loads to rely in part on balancing energy to provide near-real-time price signals and foster demand response. Under the relaxed balanced schedule policy, a load-serving entity could elect to purchase a share of its generation requirements from the balancing energy market. While this leaves the REP un-hedged and exposed to price fluctuations in the balancing energy market, many REPs and large energy consumers found this strategy advantageous since balancing energy prices tend to be lower than the average cost of firm generation obtained from bilateral contracts, particularly if they served loads with some capability to reduce energy usage in the face of high market prices.

Finally, some industrial energy consumers participate in curtailment programs that are established by REPs. These programs are conducted outside of the formal ERCOT market and are used by REPs to shape their generation needs and reduce their costs. The absence of regulatory oversight permits REPs to negotiate and customize these arrangements to match the consumers' needs and flexibility.

If an energy consumer opts to offer its interruption capability into an ancillary service market, then its ability to react to wholesale balancing energy prices, avoid the four summer peaks, and participate in any REP-sponsored demand response programs will be constrained. If the load is providing responsive reserves, then ERCOT monitors the load's level every 3 s to ensure that the load is available for interruption should the system need to rely upon the interruption to maintain frequency. A QSE could incur a penalty (a scheduling control error) if it is not providing its committed level of operating reserves. Thus many of ERCOT's most flexible, interruptible, or potentially price elastic electric loads will not react to prices. This appears to be most true for LaaRs which have a bilateral contract to provide responsive reserves to a REP/QSE and are self-scheduled. Two-thirds of the LaaRs providing ancillary services are self-arranged by a REP or QSE through bilateral contracts in this manner.

The amount of load reduction from voluntary load response actions during a spike in balancing energy prices or a summer peak

Table 3
Categorization of LaaRs by size[a].

LaaR capacity range	Number of LaaRs	Total capacity (MW)
1–10 MW	113	428
11–50 MW	25	487
51–100 MW	1	66
Greater than 100 MW	5	1003

[a] Source: information provided by the ERCOT staff to the author on January 21, 2009.

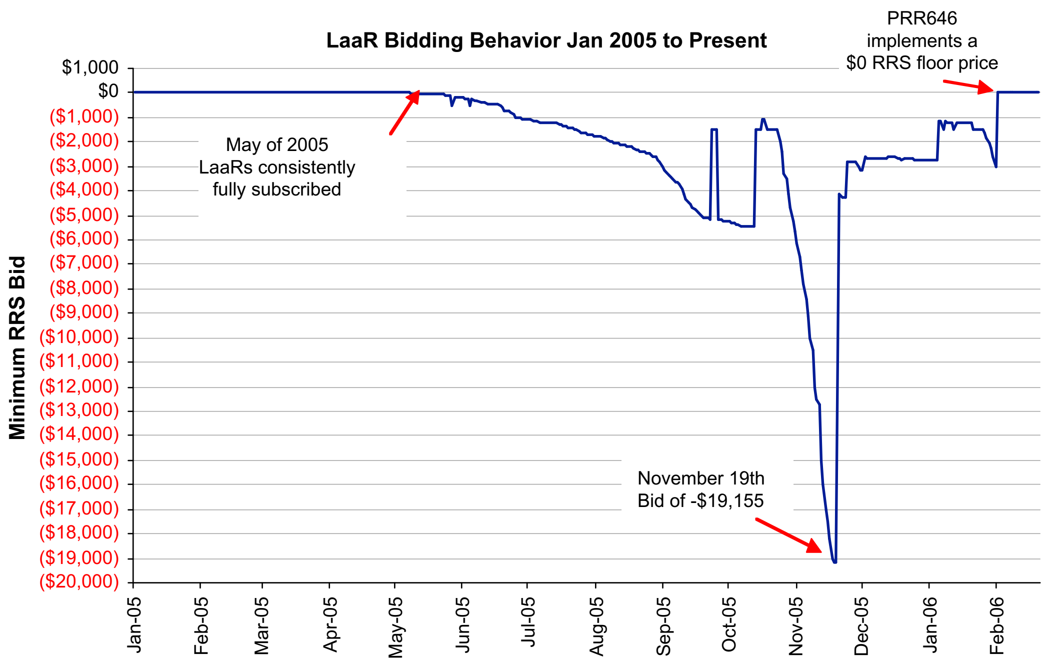

Fig. 1. LaaR Bidding Behavior Prior to the Introduction of a Floor on Prices which may be offered.

is thought to be roughly 600 MW. Based on some simple comparisons of the aggregate load levels of transmission voltage (large industrial) energy consumers between days of likely 4 CP charges and adjacent days, the ERCOT staff has identified about 600 MW of aggregate demand response, or about a 1% reduction in demand [12]. The average own-price elasticity for the aggregated block of all energy consumers in ERCOT with IDRs is about −0.000008 [13], which could imply a load reduction of about 600 MW during a spike in balancing energy prices. A survey of load-serving entities conducted by the ERCOT staff in 2007 found that 184 MW of load is capable of shifting in response to time-of-use pricing, 91 MW of load responds to critical peak pricing, and 431 MW of load responds to real-time pricing price signals (with the majority of this load with strike prices below $300 per MWh) [14]. 222 MW of load responds to 4 CPs, although much of this load may also be served under time-of-use or real-time pricing arrangements. Thus, each of these calculations supports an estimate of about 600 MW of demand reduction during a price spike.

Relying upon only 1% of the total demand side of the market to respond to prices does not provide sufficient demand response to constrain prices to reasonable levels, thus necessitating wholesale offer caps. Yet, a small amount of demand response can still have a tremendous impact on wholesale market prices. Like many other markets [15], the wholesale supply curve or bid stack becomes nearly vertical beyond some point. It is quite possible for a modest amount of demand reduction (100 MW, for example) to yield a reduction of $1000 per MWh in balancing energy market prices during a price spike. Attempts to further increase demand response to price signals include the BUL program and efforts to extend demand response opportunities to residential and small commercial energy consumers.

5. BULs

A second avenue to facilitate the response of industrial energy consumers to price signals was established. "Demand bidding" is permitted, whereby consumers can submit a formal offer to the balancing energy market describing a strike price at which they would curtail and a curtailment amount.[4] Under the BUL program, a load that submits an offer and is struck can also receive a capacity payment based on the prevailing price of non-spinning reserves. This capacity payment was introduced to recognize that a response to wholesale prices which is visible to the market is more valuable than simple voluntary load response which is not visible to the market and cannot be relied upon by ERCOT's system operators in matching supply with demand. BUL offers may be made just a few minutes in advance of each 15-min settlement interval and after the price of non-spinning reserves is already established in ERCOT's day-ahead market for ancillary services, so that this additional incentive is known to the BUL at the time the BUL submits its offer to reduce its energy purchases.

However, the capacity payment (which is normally zero) has not been sufficient to induce consumers into submitting formal offers. Rules requiring all BULs under a QSE's control to be scheduled as a group and complicated baseline formulas (modeled after an early version of those developed for the New York ISO's Emergency Curtailment Program) have also been cited as impediments to program participation [16]. Since energy consumers can receive most of the benefits associated with price response by voluntary or passive load response – without the hassles associated with making formal offers to ERCOT's balancing energy market and without the potential penalties which would be incurred if the load failed to achieve its promised level of demand response – this initiative has not succeeded.

6. Demand response involving smaller energy consumers

The programs and opportunities identified above exclusively involve industrial energy consumers, equipped with the metering equipment necessary to accurately measure and confirm their response to a curtailment request or price signal. Once options for demand-side participation by industrial energy consumers were

[4] Note that a formal day-ahead market will not be established in ERCOT until 2010 or 2011.

opened, the focus switched toward providing smaller loads with opportunities for demand-side participation.

Prior to restructuring, Texas had achieved some success in extending demand response opportunities to smaller energy consumers. Houston Lighting and Power Company (HL&P) operated a large load control program which controlled residential air conditioners. Central and Southwest Corporation's Customer Choice and Control pilot program in Laredo in the mid-1980s was an early experiment in two-way communications and real-time pricing involving residential consumers. However, demand response programs involving smaller loads failed to survive the transition to a restructured market.

The implementation of load control programs was declared to be a competitive energy service, and therefore best left to the competitive market. Consequently, the HL&P regulated "wires company" (presently known as CenterPoint Energy Houston) was required to terminate the direct load control program that it operated prior to restructuring. The program assets were transferred to Comverge, who was unable to continue the program due to difficulties inherent in dealing with the many REPs who sold power to the program participants, extensive measurement and verification requirements placed upon the program by the ERCOT stakeholders and the PUCT Staff, and problems in securing any compensation from the REPs and other market participants who were likely to realize a benefit from interruptions.

Some of the larger REPs showed renewed interest in residential load control in 2008, but were unable to convince the PUCT to endorse a deemed savings approach to the quantification of savings for ERCOT's wholesale market settlement in lieu of a more-costly quantification approach requiring meters on a sample of program participants.

There is hope that many of the problems inherent in operating residential demand response programs in the areas opened to competition will be overcome with the completion of advanced metering infrastructure (AMI) in the Oncor and CenterPoint service areas. With the new AMI and broadband over power line (BPL) and power line carrier (PLC) systems under development in the two largest service areas opened to retail competition, consumption can be monitored at 15-min or 30-min intervals, enabling REPs and transmission and distribution utilities to monitor consumer response to curtailment requests or price signals. It is anticipated that ERCOT's settlement systems will eventually rely upon actual metered load data for small consumers, rather than the statistical load profiles used today. By providing a two-way communication infrastructure, participants in a residential or small commercial demand response program could agree to place water heaters, air conditioners, and other equipment under the control of a REP or program administrator. A pilot program directly involving three REPs, two transmission and distribution utilities, and a number of technology providers under the administration of the Center for the Commercialization of Electric Technologies demonstrated the capabilities of the new systems in 2008.

While the establishment of a better metering and communication infrastructure should provide the needed platform for demand response involving smaller energy consumers, other challenges will remain. Until ERCOT's systems can process the information collected by the new AMI systems, expensive measurement and verification activities will be required in order to enable ERCOT to recognize residential demand response in its settlement systems. REPs may be reluctant to make long-term investments in control equipment at customer premises in the absence of a long-term commitment from the customer. Coordination among REPs, transmission and distribution utilities (who own and operate the metering and communications systems), ERCOT, and various technology providers is cumbersome. Until all of the transmission and distribution utilities in ERCOT's competitive areas invest in AMI systems, different opportunities to implement residential demand response programs will be present in different locations within the market.

7. Emergency interruptible load service

After many years of discussion and proposals, a new Emergency Interruptible Load Service (EILS) program was finally launched in mid-2008. EILS is roughly modeled after the emergency curtailment programs established in other wholesale markets in the northeast U.S. and California.

For years, there was great reluctance to establish any "demand only" program outside of the market structure, since it might be perceived as an admission that market forces alone could not provide adequate demand response or as some indication of a market failure. However, a number of factors made an out-of-market demand response program critical. In 2006, the PUCT voted to pursue an energy-only resource adequacy approach and permitted higher caps on wholesale prices, thus placing greater reliance on demand response to mitigate unreasonable price spikes. During a system emergency on April 17, 2006, there were industrial loads that were ready, willing, and able to be interrupted to restore reliability, but ERCOT had no mechanism to interrupt them. Many interruptible loads were not providing an ancillary service at the time, and there was no program through which curtailments could be requested. A large "waiting list" of industrial interruptible loads which were willing and able to participate in the crowded ancillary services markets had developed. Finally, actual planning reserves dipped below 3% during the summer of 2006. The PUCT adopted ERCOT's proposal to establish EILS in 2007 after rejecting an alternative proposal from the Steel Mill Coalition [17].

Under EILS, interruptible loads which are not otherwise providing an operating reserve receive a bid-based payment for curtailing their purchases of electricity from the grid when ERCOT is in an emergency. Program participants in the EILS must curtail within 30 min notice if the second level of a system emergency was declared [18].

8. Demand response in the future nodal market

When the ERCOT market transitions to a nodal design toward the end of 2010, it is likely to take "one step forward and two steps back." The new day-ahead energy market may open up opportunities for demand response by industrial loads with predictable and flexible energy needs. However, opportunities for loads to respond to real-time prices will become limited. Testifying on behalf of the staff of the PUCT, a witness explained the concern over permitting demand response in real-time:

> Passive demand response occurs when loads reduce their consumption in response to prices they observe without actively submitting price-sensitive offers into the wholesale market. The prices that are posted are based on the load levels that ERCOT observes and the generation and load resource offers it has received from suppliers. If ERCOT posts a price 10 min in advance and loads respond, then the price for that interval is incorrect. For example, imagine on a high load day, ERCOT posts a price of $200 per MWh 10 min in advance. Further, assume that observing this price, load reduces its consumption by 2000 MW. Finally, assume that the price would have been $100 per MWh had the 2000 MW not been included in the dispatch. In this case, dispatch signals will cause generators to over-generate by 2000 MW. In addition, ERCOT will not be revenue-neutral. Generators will have received payments for 2000 MW

of generation that will not be collected from the loads, since loads are not consuming that amount. Hence, an uplift charge will be needed to collect the additional revenue needed to compensate the generation [19].

For these reasons, the PUCT has sent mixed signals regarding its interest in demand response in real-time as it refines its wholesale market structure. To reduce forecasting error and to discourage loads from responding to prices in real-time, advanced notice of prices will be eliminated. Also, a penalty (the Reliability Unit Commitment Capacity Short Charge) will be used to discourage REPs (and their customers) from relying upon real-time purchases (as opposed to bilateral contracts and the forthcoming day-ahead market) to secure generation. Further complications arise from the use of zonal prices to settle energy purchases, while nodal prices are used to establish the value of a supply-side resource in the market. While taking these steps to discourage price-chasing may provide system operators with better demand forecasts, some demand response in real-time is likely to be sacrificed.

As the PUCT approved "nodal protocols" which included the features mentioned above to discourage demand response, it nonetheless expressed interest in exploring new avenues for demand response. Under one proposal – consistent with priority pricing [20] – loads would provide the ISO with a commitment to curtail at certain price points, in return for protection from various penalties and other incentives [21].

9. Observations

By any reasonable estimate, the amount of demand response in ERCOT has not yet regained its pre-restructuring levels. The amount of interruptible load available to be dispatched by a utility prior to restructuring is compared to the amounts of load qualified as LaaRs in Fig. 2. Over 3000 MW of interruptible load was available in ERCOT prior to restructuring (not counting curtailment programs and

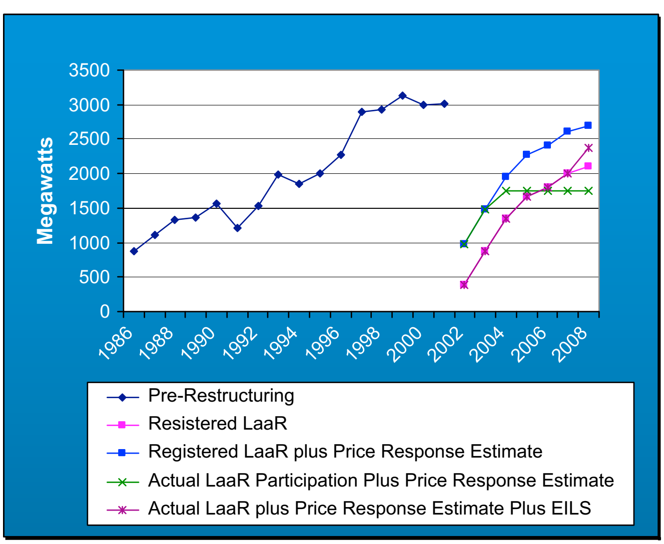

Fig. 2. Quantity of demand-side resources in ERCOT over time. LaaR offers to provide Response Reserves. Sources: data for 1985–1993 is from PUCT 1996 Statewide Electrical Energy Plan for Texas, June 1996, and represents interruptible loads plus a small contribution from various load cycling programs. Interruptible load data for 1994–1999 is from PUCT Project 22209 Annual Update of Generating Electric Utility Data, 2000. Data for 2000 and 2001 are from ERCOT capacity, demand and reserve reports for those years. LaaR data are from the ERCOT staff. Actual LaaR participation data reflect the 1150 MW cap on LaaR provision of responsive reserves. EILS estimate reflects average amounts procured (i.e., 275 MW) in recent solicitations for this resource; 600 MW is used as the estimate of price response.

residential direct load control). This amount dropped to almost nothing after interruptible tariffs were terminated. The amount of load qualified as LaaR is presently nearly 2200 MW. However, the cap on LaaR participation providing responsive reserves of 1150 MW tends to reduce the quantity of demand-side resources that the ISO can rely upon to provide an ancillary service or interrupt in response to an emergency condition at any given time.[5] Even including price-responsive load (which is normally not counted as a resource given its lack of commitment to deploy during an emergency), the demand side of the market does not provide the same level of demand response as was provided prior to restructuring. However, the growth in the development of the demand side provides reason for optimism. If this growth continues (as it should, once demand response opportunities are better provided to smaller consumers),[6] over 3000 MW of demand response may be again available to the market within a few years (Fig. 2).

Should ERCOT's market-based approach of promoting demand-side resources be replaced with a return to simple interruptible tariffs, real-time pricing tariffs, and load management programs? Certainly a well-designed simple tariff with sufficiently high-price discounts can elicit considerable participation. Yet, a return to simple tariffs is unlikely. The pre-restructuring interruptible tariffs were often criticized for providing too high a price discount when generation was ample and too low an incentive for participation when reserve margins were inadequate.

A credible estimate of the benefits associated with demand response in the ERCOT market has not yet been produced. However, the benefits are no doubt enormous. As in many other wholesale electricity markets in North America, the supply curve for balancing energy is nearly vertical at relatively high levels of demand. Consequently, even a modest amount of voluntary load response can have a considerable impact on market prices.[7] When demand-side participation in ancillary services markets declined in late 2008 in response to Hurricane Ike and the national recession, prices in ancillary service markets increased, even though natural gas prices had declined.

10. Conclusions

In some respects, ERCOT's experience is not unique. While there has likely been a general decline in dispatchable demand-side resources in North America over the past decade, this decline is most pronounced in the markets which have achieved the greatest degree of restructuring [23]. By Federal Energy Regulatory Commission's (FERC) calculation, ERCOT went from having one of the nation's largest demand side bases prior to restructuring to the smallest overall "existing demand response resource contribution" of any market in the U.S., at about 3% of peak demand. [23] However, FERC's simple comparisons may be misleading, since there are many forms demand response may take in a restructured market such as ERCOT which may not readily fit into FERC's category of "load management." And a simple MW metric might not be best.

A favorable market design is critical if there is a desire to value and accommodate demand response within the market framework. But even in a restructured market, regulatory intervention may be necessary. Regulatory oversight over the market design process may be needed in order to establish a market environment which

5 In addition a small amount of load provides regulation service and non--spinning reserves.

6 A curious dip in LaaR and EILS participation became evident in late 2008. This drop might be associated with the impacts of Hurricane Ike or the national recession.

7 A 250 MW reduction in demand can readily produce a 10-fold decrease in the market-clearing price of energy at high levels of demand [22].

will facilitate demand response. Even with a friendly market design, there may be a need to establish emergency programs outside of a market framework.

An advantage of a competitive market is that it permits negotiation between REPs and consumers, which can in turn introduce some creativity in pricing and greater flexibility in the types of arrangements. Greater customization provides benefits. Presumably, the resulting negotiations result in pricing which is in line with the true value of any demand-side resource (under the prevailing market design).

Clearly, a lot of work remains to ensure that North America's most successful restructured market provides adequate levels of demand-side participation. Yet, some optimism may be in order. The potential for greater demand-side participation in electricity markets is greater than ever. Widespread deployment of smart metering systems will enable demand response by smaller energy consumers. Technology will advance. Hopefully, our understanding of consumer behavior and the policies and market structure necessary to enable demand response will similarly improve.

References

[1] Alliance for Retail Choice. ARC's baseline assessment of choice in the United States (ABACUS), http://www.allianceforretailchoice.com/ABACUSpublication.pdf; May 2007.
[2] Center for Advancement of Energy Markets. Retail energy deregulation index. 4th ed. Center for Advancement of Energy Markets, http://www.caem.org/website/pdf/RED2003.pdf; 2003.
[3] Federal Energy Regulatory Commission. Working paper on standardized transmission service and wholesale electric market design, http://www.ferc.gov/docs-filing/elibrary.asp; March 15, 2002.
[4] Adib P, Zarnikau Jay. Texas: the most robust electricity market in North America. In: Sioshansi F, editor. Electricity market reform: an international perspective. Elsevier; 2006. p. 383–418.
[5] Zarnikau J. Using interruptible load as an ancillary service in the restructured ERCOT market. US Energy Association Dialogue. Available from:, http://www.usaee.org/pdf/Aug06.pdf#13d; 2006, July.
[6] PUCT Market Oversight Division. Project No. 22209: 2000 annual update of generating electric utility data, http://www.puc.state.tx.us/; December 2000. Austin, TX.
[7] Public Utility Commission of Texas. Final order in docket No. 23220: petition of the Electric Reliability Council of Texas for approval of the ERCOT protocols, http://www.puc.state.tx.us/; 2000. Austin, TX.
[8] Public Utility Commission of Texas. Docket No. 31540: proceeding to consider protocols to implement a Nodal market in ERCOT pursuant to PUC Subst. R. 25.501, http://www.puc.state.tx.us/; November 2005. Austin, TX.
[9] Electric Reliability Council of Texas. ERCOT protocols, www.ercot.com.
[10] ERCOT Staff for the Reliability and Operations Subcommittee. Utilizing high-set load shedding schemes to provide responsive reserve services, http://www.ercot.com; November 2002. Austin, TX.
[11] ERCOT Demand Side Working Group. Load participation in the ERCOT Nodal market, http://www.ercot.com; 2007.
[12] Jones Sam, Wattles Paul, Krein Steve. ERCOT emergency load response. PUCT demand response workshop. See, http://www.puc.state.tx.us/electric/projects/32853/091506/ERCOT.pdf; September 15, 2006.
[13] Zarnikau J, Hallett Ian. Aggregate industrial energy consumer response to wholesale prices in the restructured Texas electricity market. Energy Economics 2008;30(4):1798–808.
[14] ERCOT Staff. Load response survey: preliminary results. Presentation to ERCOT's Demand Side Working Group; August 10, 2007.
[15] Caves D, Eakin Kelly, Faruqui Ahmad. Mitigating price spikes in wholesale markets through market-based pricing in retail markets. Electricity Journal 2000;13(3):13–23.
[16] Trefny F. BUL program status. Presentation at PUCT demand response workshop. See also: http://www.puc.state.tx.us/electric/projects/32853/32853.cfm; December 8, 2006.
[17] Public Utility Commission of Texas. Docket No. 32615: petition of Chaparral Steel Co., Structural Metals, Inc., and Nucor steel for the adoption of a rule to implement a tiered frequency response service, http://www.puc.state.tx.us/.
[18] Public Utility Commission of Texas. Docket No. 32853: evaluation of demand response programs in the competitive electric market, http://www.puc.state.tx.us/.
[19] Patton D. Rebuttal testimony on behalf of the Staff of the Public Utility Commission of Texas, PUCT Docket No. 31540: proceeding to consider protocols to implement a Nodal market in ERCOT pursuant to PUC Subst. R. 25.501, http://www.puc.state.tx.us/; November 2005.
[20] Oren S, Stephen Smith, Robert Wilson. Priority service: unbundling quality attributes of electric power. EPRI Report EA-4871; November 1986.
[21] Zarnikau, J, Shmuel Oren. Long-term BUL offers. Presentation to the ERCOT Demand Side Working Group; October 2005.
[22] Frontier Associates. Demand response pilot: status and preliminary recommendations. Presentation to the Center of Commercialization of Electric Technologies; January 9, 2009.
[23] Federal Energy Regulatory Commission. Assessment of demand response and advanced metering, http://www.ferc.gov/legal/staff-reports/demand-response.pdf; August 2006.