Load Zone	2010 – Zonal	2011 – Nodal
Houston	17.8%	14.0%
South	17.1	14.5
North	17.7	13.1
West	18.5	17.1

Table 1: Price Change as a Percent of Average Price

In well functioning markets we expect to observe a close correlation between price and load levels. This relationship was not observed under the zonal market design and was described repeatedly in prior annual reports⁵.



The relationship between average prices and average load levels during selected hours of the summer months are shown in Figure 13 and Figure 14. The periods shown in these two figures are times when there are typically large changes in load levels and associated generation ramping.

⁵ See 2009 ERCOT SOM Report at 21-28, 2008 ERCOT SOM Report at 21-28, and 2007 ERCOT SOM Report at 60-65.



Figure 14: Price Load Relationship during Summer Ramping Down Hours

The correlation between price and load is very high during the ramping up hours. This is the expected result when price formation is based on the cost of supply to meet the entire demand, rather than to meet a delta between total load and schedules, which was the case in the zonal market. The relationship between price and load during the ramping down hours exhibits discontinuities at 10:00 pm and just after midnight. These short term price increases are typically the result of prices rising in response to transitory generating unit ramp rate limitations in the aftermath of units turning off overnight. Even so, these price movements are much smaller, and less frequent than what was routinely observed in the zonal market.

D. February Cold Weather Event

A significant operational challenge greeted the nascent nodal market in the early morning of February 2nd. The resulting market outcomes had a sizable effect on the overall annual results. This section more fully describes the specifics of that event.

In the early morning hours of February 2, 2011, the ERCOT region experienced extreme cold weather conditions, record electricity demand levels, and the loss of numerous electric generating facilities across the ERCOT region. These events combined to result in the deployment of load resources contracted to provide responsive reserve service and Emergency

Interruptible Load Service ("EILS") and culminated with 4,000 MW of firm load being shed for several hours.

Shown in Figure 15 are the five days through February 2, 2011 with the highest ERCOT electricity demand at the time just prior to the deployment of load resources. The demand for electricity in the early morning of February 2nd was 2,760 MW higher than on any other day in the history of the of the ERCOT region at this same time, and was experiencing a rapid rate of growth as is typical on such cold winter mornings. The demand curve for February 2, 2011 is noticeably distorted after 5:20 a.m. due to the various stages of load shedding that started at that time and remained in effect until just after 1:00 p.m., with the exception of approximately 470 MW of EILS deployments that remained in effect until approximately 10 a.m. on February 3rd.



Also shown in Figure 15 is the estimated load that would have materialized on February 2nd absent any load curtailments, which indicates that, absent curtailments, the demand in the

ERCOT region would have approached 59,000 MW just after 7 a.m. This is almost 2,300 MW higher than the previous record instantaneous demand for electricity at this time of the day.

To provide additional perspective on the capacity limitations experienced on February 2nd, Figure 16 shows the available capacity (online capacity plus offline non-spinning reserves) and the ERCOT load for the seven days from January 31 through February 6, 2011.



Figure 16: Seven Day View of ERCOT Available Capacity and Load

The data in Figure 16 highlight the highly unusual and extremely narrow gap between available capacity and actual load that was experienced on the morning of February 2, 2011 relative to other days of similar and much lower load levels. These data also highlight the successful efforts to return substantial generating capacity to service prior to the record peak demand on the evening of February 2nd and to sustain the availability of that capacity for the high electricity demands experienced again on February 3rd.

As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal cost of the marginal action is that associated with the dispatch of the last

generator required to meet demand. It is appropriate and efficient in these hours for this generator to "set the price." However, this is not true under shortage conditions. When the system is in shortage, the demand for energy and operating reserves cannot be satisfied with the available resources, which will cause the system operator to take one or more of the following actions:

- Sacrifice a portion of the operating reserves by dispatching them for energy;
- Voluntarily curtail load through emergency demand response programs;
- Curtail exports or make emergency imports; or
- Involuntarily curtail load.

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When the market is in shortage, the marginal action taken by the system operator is generally to not satisfy operating reserves requirements (*i.e.*, dispatching reserves for energy). Diminished operating reserves results in diminished reliability, which has a real cost to electricity consumers. In this case, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission.

During the morning of February 2, 2011, ERCOT operating reserve levels were reduced to perilously low levels for a sustained period of time. ERCOT's primary measure of overall operating reserves is Physical Responsive Reserve ("PRC"). ERCOT will remain in various levels of Energy Emergency Alert ("EEA") once PRC drops below 2,300 MW. Figure 17 shows the wholesale market prices and PRC from 21:00 on February 1 through 21:00 on February 2, 2011.

The data in Figure 17 show increased price volatility from 3:30 to 4:45 a.m. as system demand was increasing and generating units continued to be in various stages of tripping and starting. By 4:55 a.m., prices had reached a sustained level \$3,000 per MWh, and PRC dropped below 2,300 MW by 5:10 a.m. PRC dropped to as low as 445 MW at 6:25 a.m., and remained consistently below the minimum 2,300 MW level until 12:00 p.m.



These wholesale market pricing outcomes were consistent with the ERCOT energy-only market design. The wholesale market prices began communicating the degradation in system reliability as early as 3:30 a.m. By 4:55 a.m. – 15 minutes prior to the reduction of PRC below the minimum acceptable level of 2,300 MW and 50 minutes prior to the first stage of firm load shedding – prices were consistently communicating the rapidly deteriorating system reliability conditions. Finally, as load levels naturally reduced and reserve levels were restored, prices dropped back to levels typical of non-shortage conditions.

The secondary effect of the conditions during the morning of February 2, 2011 was the effect on the day-ahead market for February 3, 2011. Figure 18 shows the hourly average day-ahead market energy prices for February 1st through the 5th. Notable in Figure 18 is that, while somewhat higher than a typical day, the day-ahead prices for February 2nd are significantly lower than the real-time prices shown in Figure 17 for the same day. Figure 18 also shows that the day-ahead prices for February 3rd were substantially higher than on February 2nd and, in fact,

represent the highest day-ahead market prices experienced since the implementation of the nodal market.



Figure 18: Average Hourly Day-Ahead Prices for Feb. 1-5, 2011

To better understand these day-ahead pricing outcomes for February 3^{rd} requires a review of the day-ahead market function and timing. The ERCOT day-ahead market is not a mandatory market; rather, it is a voluntary market that consists of willing sellers that will be cleared for offers to sell energy at their offer price or higher and willing buyers that will be cleared for bids to buy at their bid price or lower. The day-ahead market is not limited to physical generation as sellers or physical load serving entities as buyers. In other words, any market participant – whether it has a physical position in the market or not – can participate in the day-ahead market and take a financial position against the real-time market. Because of the voluntary, financial nature of the day-ahead market, its outcomes are strongly driven by expectations of the real-time market performance for the following day.

On this point, an understanding of the timing of the day-ahead market execution is critical. The day-ahead market opens for bid/offer submission at 6:00 a.m. on the day prior to the operating day, and the submission window closes at 10:00 a.m. Thus, for the February 3rd day-ahead

market, the submission window opened at 6:00 a.m. and closed at 10:00 a.m. on February 2nd. Thus, at the time that bids/offers were submitted for the February 3rd day-ahead market, ERCOT was in the middle of the EEA level 3 events on February 2nd. Considerable uncertainty regarding generating unit availability and system conditions for February 3rd existed at that time, while the forecast called for continued arctic conditions across the state and record electricity demand was again forecast for the ERCOT region.

On a typical day, the day-ahead market results for February 3rd would give rise to market performance concerns, just as the real-time results on February 2nd would also raise concerns on a typical day. However, the real-time system conditions on February 2nd were far from typical. with the market outcomes reflecting the underlying system reliability conditions, consistent with the energy-only market design. Likewise, the day-ahead market outcomes for February 3rd were driven by the highly atypical uncertainties and risks facing both the supply and demand sides that existed at the time the day-ahead market submissions occurred, and the results are not unexpected given those considerations. Notably, while the day-ahead prices for February 3rd averaged \$465.64 per MWh, day-ahead prices for February 4th and 5th averaged \$99,56 and \$44.68 per MWh, respectively, as the weather moderated resulting in decreased electricity demands and generation resources previously experiencing outages were returned to service. Although near-record electricity demand levels were again experienced on February 3rd, a substantial number of generating units that were forced out of service on February 2nd were able to return by the morning of February 3rd. Real-time prices on February 3rd averaged approximately \$112 per MWh, which is higher than a typical day but much lower than the dayahead prices for that day. Overall, we find that the real-time and day-ahead wholesale markets for February 2nd and 3rd operated efficiently given the system conditions and the outcomes are consistent with the ERCOT energy-only wholesale market design.

Although a wide range of actions were undertaken by generation resource owners in preparation for the extreme weather conditions, it is clear from the unprecedented loss of generation capacity on the morning of February 2nd that many of these preparatory efforts were unsuccessful. This experience will serve to produce lessons learned and specific areas for improvement in the areas of generation resource weatherization and coordinated extreme weather planning. Overall, although the scope and magnitude of the generating unit outages on February 2nd was absolutely

unprecedented, we do not find any evidence that indicates that any of the outages were the result of physical withholding.

Another measure to provide additional insight related to this finding is the relative profitability of market participants during these events and how it correlates with unit outages. Although an assessment of profitability in isolation is insufficient to draw conclusions related to market manipulation or market power, increased profitability is the primary motive associated with resource withholding strategies. Hence, a negative correlation between resource outages and profitability would provide increased confidence in the finding that the outages were not the result of market manipulation strategies or market power abuses.



Figure 19: Generation Availability and Net Financial Position on Feb. 2, 2011

Real-time market prices on the morning of February 2nd were at or near the system-wide cap of \$3,000 per MWh due to the short-supply conditions existing during the EEA event. Figure 19 shows the relationship between wholesale market profitability on February 2nd and availability of generation during the morning of February 2nd for market participants representing the largest

fleets of generating resources. The data in Figure 19 show that those market participants who were able to operate their generation fleet at greater than 90% availability during the morning of February 2nd were financially successful that day. In contrast, market participants affected by significant generation outages found themselves unprofitable that day.⁶

Day-ahead market prices for February 3rd were also affected by the conditions on February 2nd and were substantially higher than normal levels. Although some market participants that lost money on February 2nd were able to recover much of their lost generating capacity and financial losses on February 3rd, none of the market participants that lost significant generating capacity and were unprofitable on February 2nd had financial gains on Feb. 3rd that significantly exceeded their losses on February 2nd.

E. August Weather Conditions and Shortages

The summer of 2011 will be remembered as the hottest and driest on record in ERCOT. These extreme weather conditions led to record high demand for electricity during August. There were 50 hours in 2011 with electricity demands that exceeded the highest hourly demand that occurred in 2010. More details of the demand for electricity in ERCOT are provided in Section IV.A, ERCOT Loads in 2011.

During these high demand conditions there is an increased likelihood that the available generation capacity is not sufficient to meet customer demands for electricity and maintain the required reliability reserves. As more fully described later in Section V.B, Effectiveness of the Scarcity Pricing Mechanism, the nodal market causes energy prices to rise toward the system-wide offer cap as available operating reserves approach minimum required levels to reflect the degradation in system reliability. Figure 20 shows the aggregated amount of time represented by all dispatch intervals where the real-time energy price was at the system-wide offer cap, displayed by month. Of the 28.5 hours of the annual total time at the system-wide offer cap, more than 17 hours (60 percent) occurred during August.

⁶ The data in Figure 19 do not include market participants without physical generation resources or market participants operating only wind generation resources or relatively small fleets of non-wind generation resources. Outage capacity does not include planned outages.



Figure 20: Duration of Prices at the System Wide Offer Cap

The next figure provides a more detailed comparison of load, required reserve levels, and prices in July and August. The weather in ERCOT was extremely hot and dry during both months, but there were very few dispatch intervals when real-time energy prices reached the system-wide offer cap in July compared to the relatively high frequency it occurred in August. Although the weather may have been similar, there were significant differences in load and available operating reserve levels, resulting in much higher prices in August.

Shown on the left side of Figure 21 is the relationship between real-time energy price and load level for each dispatch interval during the months of July and August. ERCOT loads were greater than 65 GW for three hours in July, whereas load levels exceeded 65 GW for 71 hours in August. As previously discussed, a strong positive correlation between higher load and higher prices is expected in a well functioning energy market. We observe such a relationship between higher prices and higher loads in both months. With overall higher loads and more frequent occurrences of very low operating reserves in August, higher energy prices are expected.

Although load levels are strong predictors of energy prices, an even more important predictor is the level of operating reserves. Simply put, operating reserves are the difference between the total capacity of operating resources and the current load level. As load level increases against a fixed quantity of operating capacity, the amount of operating reserves diminishes. The minimum required operating reserves prior to the declaration of Energy Emergency Alert Level 1 by ERCOT is 2,300 MW. As the available operating reserves approach the minimum required amount, energy prices should rise toward the system-wide offer cap to reflect the degradation in system reliability.





On the right side of Figure 21 are data showing the relationship between real-time energy prices and the quantity of available operating reserves for each dispatch interval during the months of July and August. This figure shows a strong correlation between diminishing operating reserves and rising prices. In July available operating reserves were generally maintained well above minimum levels, and there were only 0.22 hours where the energy price reached \$3,000 per MWh. In contrast, there were numerous dispatch intervals in August when the minimum

operating reserve level was approached or breached, with 17.4 hours where prices reached \$3,000 per MWh. However, there are also a substantial number of dispatch intervals where operating reserves are below minimum requirements with prices well below the level that would be reflective of the reduced state of reliability. In Section V.C, Demand Response Capability we provide an example explaining why this can occur and offer a recommendation for improvement.

II. REVIEW OF DAY-AHEAD MARKET OUTCOMES

One of the fundamental improvements brought about by the implementation of ERCOT's nodal market design is the establishment of a centralized day-ahead market, which allows participants to make financially binding forward purchases and sales of power for delivery in real-time. Offers to sell can take the form of either a three-part supply offer, which allow sellers to reflect the unique financial and operational characteristics of a specific generation resource, or an energy only offer, which is location specific but is not associated with a generation resource. Bids to buy are also location specific. Ancillary services are also procured as part of the day-ahead market clearing. The third type of transaction included in the day-ahead market is bids to buy point to point ("PTP") Obligations, which allow parties to hedge the incremental cost of congestion between day-ahead and real-time operations.

With the exception of the acquisition of ancillary service capacity, the day-ahead market is a financial market. Although all bids and offers are evaluated in the context of their ability to reliably flow on the transmission network, there are no operational obligations resulting from the day-ahead market clearing. These transactions are made for a variety of reasons, including satisfying the participant's own supply, managing risk by hedging the participant's exposure to the real-time market, or arbitraging with the real-time markets. For example, load serving entities can insure against volatility in the real-time market by purchasing in the day-ahead market. Finally, the day-ahead market plays a critical role in coordinating generator commitments. For all of these reasons, the performance of the day-ahead market is essential.

In this section we review energy pricing outcomes from the day-ahead market and compare their convergence with real-time energy prices. We will also review the volume of activity in the day-ahead market, including a discussion of Point to Point Obligations. We conclude this section with a review of the ancillary service markets.

A. Day-Ahead Market Prices

One indicator of market performance is the extent to which forward and real-time spot prices converge over time. Forward prices will converge with real-time prices when two main conditions are in place: a) there are low barriers to shifting purchases and sales between the forward and real-time markets; and b) sufficient information is available to market participants to allow them to develop accurate expectations of future real-time prices. When these conditions are met, market participants can be expected to arbitrage predictable differences between forward prices and real-time spot prices by increasing net purchases in the lower-priced market and increasing net sales in the higher-priced market. These actions will tend to improve the convergence of forward and real-time prices.

In this section, we evaluate the price convergence between the day-ahead and real-time markets. This analysis reveals whether persistent and predictable differences exist between day-ahead and real-time prices, which participants should arbitrage over the long-term. To measure the short-term deviations between real-time and day-ahead prices, we also calculate the average of the absolute value of the difference between the day-ahead and real-time price on a daily basis.

This measure captures the volatility of the daily price differences, which may be large even if the day-ahead and real-time energy prices are the same on average. For instance, if day-ahead prices are \$70 per MWh on two consecutive days while real-time prices are \$40 per MWh and \$100 per MWh on the two days, the absolute price difference between the day-ahead market and the real-time market would be \$30 per MWh on both days, while the difference in average prices would be \$0 per MWh.

Day-ahead prices averaged \$46 per MWh in 2011 compared to an average of \$43 per MWh for real-time prices.⁷ This slight premium is consistent with expectations due to the much higher volatility of real-time prices. Risk is lower for loads purchasing in the day-ahead and higher for generators selling day-ahead. The higher risk for generators is associated with the potential of having a forced outage and buying back energy at real-time prices. This may explain why the highest premiums occurred during the highest priced months. Overall, the day-ahead premiums were very similar to the differences observed in 2009 and 2010.⁸ The average absolute difference between day-ahead and real-time prices was \$24.50 per MWh in 2011; much higher than in the previous two years where the average absolute difference was \$12.25 and \$12.37 in 2010 and 2009, respectively. This large increase was the result of the significant periods of very

⁷ These values are simple averages, rather than load-weighted averages presented in Figure 1 and Figure 2.

⁸ In 2009 and 2010 under the zonal market the comparison was made between on-peak forward prices and prices for the same on-peak period in the balancing energy market.

high real-time prices during February and August. Removing the contribution from these two months reduces the average absolute difference to \$11.49 per MWh in 2011.

Figure 22 shows the price convergence between the day-ahead and real-time market, summarized by month.



Figure 22: Convergence between Forward and Real-Time Energy Prices

Below, in Figure 23 monthly day-ahead and real-time prices are shown for each of the geographic load zones. Of note is the difference in the west zone data compared to the other regions. The higher volatility in west zone pricing is likely associated with the uncertainty of forecasting wind generation output and the resulting price levels between day-ahead and real-time.



Figure 23: Day-Ahead and Real-Time Prices by Zone

B. Day-Ahead Market Volumes

Our next analysis summarizes the volume of day-ahead market activity by month. In Figure 24, we find that day-ahead purchases are approximately 40 percent of real-time load. These energy purchases are met through a combination of generator specific and virtual offers.





By adding in the effects of net energy flows associated with purchases of PTP Obligations, we find that on average total volumes transacted in the day-ahead market are greater than real-time load.

Figure 25 presents the same day-ahead market activity data summarized by hour of the day. In this figure the volume of day-ahead market transactions is disproportionate with load levels between the hours of 6 and 22. Since these times align with common bilateral transaction terms, it appears that market participants are using the day-ahead market to trade around those positions.



Figure 25: Volume of Day-Ahead Market Activity by Hour

C. Point to Point Obligations

Purchases of Point to Point ("PTP) Obligations comprise a significant portion of day-ahead market activity. These instruments are similar to, and can be used to complement Congestion Revenue Rights (CRRs). CRRs, as more fully described in Section III.C, are acquired via monthly and annual auctions and allocations. CRRs accrue value to their owner based on locational price differences as determined by the day-ahead market. PTP Obligations are instruments that are purchased as part of the day-ahead market and accrue value to their owner based on real-time locational price differences. By acquiring a PTP Obligation using the proceeds due to them from their CRR holding, a market participant may be described as rolling their hedge to real-time.

In this subsection we provide additional details about the volume and profitability of these PTP Obligations.



Figure 26: PTP Obligation Volume

Figure 26 presents the total volume of PTP Obligation purchases divided into three categories. Compared to the previous two figures which showed net flows associated with PTP Obligations, in this figure we examine the total volume. For all PTP Obligations that source at a generator location, we attribute capacity up to the actual generator output as a generator hedge. From the figure above we see that this comprised most of the volume of PTP Obligations purchased. The remaining volumes of PTP Obligations are not directly linked to a physical position and are assumed to be purchased primarily to profit from anticipated price differences between two locations. We further separate this arbitrage activity by type of market participant. Physical parties are those that have actual real-time load or generation, whereas financial parties have neither. We find that the arbitrage activity is fairly evenly split between physical and financial parties, and further, the volume of arbitrage activity steadily increased throughout the year.

To the extent the price difference between the source and sink of a PTP Obligation is greater in real-time than it was day-ahead price, the owner will have made a profit. Conversely, if the price difference does not materialize in real-time, the PTP Obligation may be unprofitable. We

compare the profitability of PTP Obligation holdings by the two types of participants in Figure 27.



Figure 27: Average Profitability of PTP Obligation

From the figure above we can infer different motivations between the two types of participants. Because financial participants have no real-time load or generation they have no other exposure to real-time prices. If a financial participant is not making a profit on their PTP Obligations there is no reason for them to buy any. In fact, their profit seeking action of buying PTP Obligations between points where congestion is expected helps make the day-ahead market converge with real-time market outcomes. On the other hand, physical participants do have exposure to realtime prices. It is reasonable to expect that this type of participant is most interested in limiting that exposure by using PTP Obligations as a hedge.



Figure 28: Point to Point Obligation Charges and Payments

To conclude our analysis of PTP Obligations, in Figure 28 we compare the total amount paid for these instruments day-ahead, with the total amount received by their holders in real-time. In most months owners received, in aggregate, more in payments for their PTP Obligations than they paid to acquire them. This occurs when real-time congestion is greater than what occurred in the day-ahead. The payments made to PTP Obligation owners come from real-time congestion rent. We assess the sufficiency of real-time congestion rent to cover both PTP Obligations and payments to owners of CRRs who elect to receive payment based on real-time prices in Section III.C, Congestion Rights Market at page 52.

D. Ancillary Services Market

The primary ancillary services are up regulation, down regulation, responsive reserves, and nonspinning reserves. Market participants may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. This section reviews the results of the ancillary services markets in 2011. In general, the purpose of responsive and non-spinning reserves is to protect the system against unforeseen contingencies (*e.g.*, unplanned generator outages, load forecast error, wind forecast error), rather than for meeting normal load fluctuations. ERCOT procures responsive reserves to ensure that the system frequency can quickly be restored to appropriate levels after a sudden, unplanned outage of generation capacity. Non-spinning reserves are provided from slower responding generation capacity, and can be deployed alone, or to restore responsive reserve capacity. Regulation reserves are capacity that responds every four seconds, either increasing or decreasing as necessary to fill the gap between energy deployments and actual system load. Figure 29 displays the quantities of each ancillary service procured each month.



Figure 29: Ancillary Service Capacity

One significant change under the nodal market is that deployments of energy occur more frequently, typically every five minutes. The more frequent deployment of energy has meant that less regulation capacity is required under the nodal market design. Even with the greater quantity of regulation capacity procured in the zonal market, ERCOT operators would resort to issuing out-of-merit instructions to one of the larger generation fleets to provide additional capacity to manage gaps between regulation and balancing energy deployments. This activity occurred 190 times during the last eleven months of the zonal market, typically occurring during periods with large changes in wind generation output. Since the outset of the nodal market there has not been a need to supplement market based deployments by calling on a single market participant, even with increased wind generation ramping.

Another change under the nodal market is that ancillary service offers are co-optimized as part of the day-ahead market clearing. This means that market participants no longer have to include their expectations of forgone energy sales in their ancillary services capacity offers. However, because clearing prices for ancillary services capacity will explicitly account for the value of energy, there is a much higher correlation between ancillary services prices and real-time energy prices. As shown in Figure 30, clearing prices for ancillary services rose quite high during February and August, corresponding with the energy prices in those months.



Figure 30: Ancillary Service Prices

In contrast to the previous data that show the individual ancillary service capacity prices, Figure 31 shows the monthly total ancillary service costs per MWh of ERCOT load and the average real-time energy price for 2008 through 2011. Figure 31 shows that total ancillary service costs are generally correlated with real-time energy price movements, which, as previously discussed, are highly correlated with natural gas price movements.

The average ancillary service cost per MWh of load increased to \$2.41 per MWh in 2011 compared to \$1.26 per MWh in 2010, an increase of 91 percent. Total ancillary service costs increased from 3.2 percent of the load-weighted average energy price in 2010 to 4.5 percent in 2011.



Ancillary Service capacity is procured as part of the day-ahead market clearing. Between the time it is procured and the time that it is required to be provided events can occur which make this capacity unavailable to the system. Transmission constraints can arise which make the capacity undeliverable. Outages or limitations at a generating unit can lead to failures to

provide. When either of these situations occurs ERCOT may open a supplemental ancillary services market (SASM) to procure replacement capacity⁹.

Figure 32 presents a summary of the frequency with which A/S capacity was not able to be provided and the number of times that a SASM was opened. The percent of time that capacity procured in the day-ahead was actually able to provide the service in the hour it was procured for was less than 20 percent of the time at the beginning of the year, increasing to more than 50 percent by the end of 2011. Even though in more than 40 percent of the hours there were deficiencies in A/S deliveries, SASMs were opened to procure replacement capacity only a fraction of the time.





In Table 2 below, we provide an annual summary of the frequency and quantity of ancillary service deficiency, where deficiency is defined as either the failure to provide or undeliverability.

⁹ ERCOT may also open a SASM if they change their ancillary service plan. This did not occur during 2011.

Service	Hours Deficient	Mean Deficiency (MW)	Median Deficiency (MW)
Responsive Reserve	4053	39	20
Up Regulation	1222	27	20
Down Regulation	1235	22	11
Non-Spin Reserve	1254	90	39

 Table 2: Ancillary Service Deficiency

Responsive Reserve service was deficient most frequently. As was the case for all ancillary services, the overwhelming majority of time, 4003 out of 4053 hours, the Responsive Reserve deficiency was due to the resource failing to provide, not because of a transmission constraint.

The next analysis, shown in Figure 33 summarizes the average quantity of each service that was procured via SASM.



Figure 33: Ancillary Service Quantities Procured in SASM

III. TRANSMISSION AND CONGESTION

One of the most important functions of any electricity market is to manage the flows of power over the transmission network by limiting additional power flows over transmission facilities when they reach their operating limits. The action taken to ensure operating limits are not violated is called congestion management. The effect of congestion management is to change generator(s) output level so as to reduce the amount of electricity flowing on the transmission line nearing its operating limit. Because the transmission system is operated such that it can withstand the unexpected outage of any element at any time, congestion management actions most often occur when a transmission element is expected to be overloaded if a particular unexpected outage (contingency) were to occur. Congestion leads to higher costs due to lower cost generation being reduced and higher priced generation increased. Different prices at different nodes are the result. With the change to ERCOT's nodal market, the decision about which generator(s) will vary their output is based on generating unit specific offer curves.

Previous annual reports described at great length the inconsistencies and resulting inefficiencies in the bifurcated way in which congestion was managed under the zonal market design. Although zonal congestion management instructions were bid-based, because all generators located within in a zone were assumed to have the same ability to affect the flows across a zonal constraint, the result was significant operational inefficiency and uncertainty. All other constraints were managed by paying generators to either increase or decrease their output from their scheduled level. Because the money to make these payments to generators was collected from all loads in ERCOT, generators had no incentive to take into account the state of the transmission system when scheduling their output.

The nodal market provides many improvements, including unit-specific offers and shift factors, simultaneous resolution of all transmission congestion, actual output instead of schedule-based dispatch, and 5-minute instead of 15-minute dispatch, among others. These changes have helped to increase the economic and reliable utilization of scarce transmission resources beyond that experienced in the zonal market, and in so doing, also dispatch the most efficient resources available to reliably serve demand.

In this section of the report we will start with a review the costs and frequency of transmission congestion in both the day-ahead and real-time markets. We will then provide a review the activity in the congestion rights market.

A. Real-Time Constraints

We begin our review by examining the real-time constraints with the highest financial impact. In all there were more than 300 different constraints active at some point during 2011. The median financial impact, as measured by congestion rent was approximately \$300,000.

Figure 34 displays the ten most highly valued real-time constraints as measured by congestion rent and indicates that the West to North interface constraint was by far the most highly valued during 2011.



Figure 34: Top Ten Real-Time Constraints

This constraint is very similar to the competitively significant constraint that existed since the inception of ERCOT's zonal market. Through the years it has been a major impediment to delivering all the wind generation located and produced in the western reaches of ERCOT to the load centers. This constraint was active at some point during every month of 2011.

Two additional constraints on the list are also related to west zone wind generation, although in different directions. The Nicole to Oak Creek constraint is a small capacity 69 kV transmission line that typically overloads under high wind conditions, while due to its load serving nature, the Odessa North 138/69 kV transformer typically overloads under low wind conditions.

The second and third constraints shown in Figure 34 are similar and reflect limitations on the amount of electricity that can be reliably imported into the Rio Grande Valley. This was most notable during the cold weather event of early February. Whereas system wide generation shortages were limited to February 2nd, extremely high customer demands for electricity coupled with the extended planned outage of local generation led to shortages and resulting load curtailments in the Valley over the next two days. Constraints limiting imports to the Valley were active and not able to be resolved for a total of 13 hours during January and February.

When a constraint becomes irresolvable, ERCOT's dispatch software can find no combination of generators to dispatch in a manner such that the flows on the transmission line(s) of concern are below where needed to operate reliably. In these situations, offers from generators are not setting locational prices. Prices are set based on predefined rules, which since there are no supply options for clearing, should reflect the value of reduced reliability for demand. In the case of these constraints related to Valley imports, the shadow price, exceeded \$4,500 per MW for the entire time they were not able to be resolved. The effects of these high prices at Valley locations were felt across the entire South Load Zone through high load zone prices.

Although the pricing outcomes were as designed, in the aftermath of this high priced congestion event, ERCOT stakeholders revisited the pricing parameters for irresolvable constraints. The outcome of this effort implemented a set of changes for pricing this constraint specifically and all irresolvable constraints in general. Specifically, due to the radial nature of the Valley Import constraint, its shadow price when irresolvable was lowered to \$2,000 per MW. To address the situation more generally, a regional peaker net margin mechanism was introduced such that once local price increases accumulate to a predefined threshold due to an irresolvable constraint; the shadow price of that constraint would drop. The IMM supported this compromise which was finalized late in the year and went into effect for 2012.

One other constraint of note shown in Figure 34 is the Garrot to Midtown 138 kV line, which limits flows within the Houston area. The transmission system upgrade necessary to resolve this longstanding, well known constraint was delayed well past its original planned in-service date. It is scheduled for completion in 2012.

The remaining constraints on the list are all fairly short duration, high impact constraints, reflecting limitations on the ability to import power to a major load center. The Carrolton Northwest to Lake Point 138 kV line and the West Denton to Jim Cristal 138 kV constraints were related to serving load in the DFW area. Singleton to Zenith limited imports to Houston from the north, while the Marion autotransformer constraint was due to an equipment deration that limited electricity flows into San Antonio.



Figure 35: Most Frequent Real-Time Constraints

Figure 35 presents a slightly different set of real-time constraints. These are the most frequently occurring. With the exception of Garrot to Midtown, described previously, all are related to wind generation. The Odessa North 138/69 kV transformer typically overloads under low wind and high local load conditions. The Lon Hill to Pelican 138 kV line was a limitation affecting a specific coastal wind generator. The other seven frequently occurring constraints are all related

to high west zone wind. Again, the West to North interface constraint tops the list as the most frequently occurring constraint in 2011. To put it in context, the West to North constraint was binding more than 20 percent of the time in 2011.

To maximize the economic use of scarce transmission capacity, the ideal outcome would be for the actual transmission line flows to reach, but to not exceed the physical limits required to maintain reliable operations. Figure 36 presents a summary of the utilization of the most active transmission constraint during 2011, the West to North interface. By comparing the actual flow with the physical limit of the constraint for each dispatch interval it was binding, we can compute its average utilization.



Figure 36: Utilization of the West to North Interface Constraint

Although there was significant variation throughout the year, the average physical limit was slightly less than 2,000 MW and the average actual flow during constrained intervals was approximately 1,500 MW. The average annual utilization of 76 percent compares favorably to 64 percent utilization experienced during the final months of the zonal market. Even more encouraging is the upward trend in utilization observed in the latter part of the year. This

increase may be attributed to increased operator confidence that generators, specifically wind generators in this case, will reduce their output as expected when the constraint is active.

There should be opportunity for increased limits in the short term and even higher utilization of this constraint as ERCOT implements more sophisticated real-time analysis of this constraint, rather than relying on off-line studies. Over the long term, the physical limit will increase as CREZ transmission projects are completed.

Although much improved, congestion management in the nodal market has not been perfect. During the spring of 2011, unexpected levels of base point and price oscillations were observed related to congestion management. The initial efforts to resolve the issue were focused on wind generation issues and managing their curtailment related to the West to North constraint. Wind generators, like all generators, are expected to continuously telemeter to ERCOT the maximum capacity output their generator can sustain ("HSL"). When wind units are not curtailed, their HSL is their current output. When wind units are curtailed, their HSL should be the maximum output they could generate if they were not being curtailed.

One of the first changes made to the nodal market systems, implemented in May 2011, was the introduction of a curtailment flag sent by ERCOT to wind generators every interval. This flag replaced the practice which required wind generators to artificially freeze their HSLs for up to 5 minutes prior to each execution of the dispatch software. Although holding the HSLs constant allowed wind generators to compare their received dispatch instructions to determine whether they had been curtailed, using stale data as input to the dispatch software created congestion management challenges.

Although the implementation of the wind curtailment flag was expected to greatly improve management of the West to North constraint, this was not the only constraint where oscillation issues were observed. Investigation by the IMM and ERCOT staff determined that oscillations were being created because ERCOT systems were using two assessments from different moments in time to calculate the generation dispatch required to resolve a constraint. The delay between the two assessments at times exceeded 5 minutes. After identifying this misalignment in constraint management calculations, ERCOT was able to resolve the issue by the end of June.

Since that time, the rare observation of constraint related oscillation has been attributable to a wind generator specific telemetry issue.

B. Day-Ahead Constraints

In this section we review transmission constraints from the day-ahead market. To the extent the model of the transmission system used for the day-ahead market matches the real-time transmission system, and assuming market participants transact in the DAM similarly to how they transact in real-time, we would expect to see the same transmission constraints appear in the day-ahead market as actually occurred during real-time.



Figure 37 presents the top ten constraints from the day-ahead market, ranked by their congestion rent. As it was in real-time, the West to North constraint has, by far, the highest financial impact. The Valley import constraint, as it was in real-time, was second on the list. The next two constraints are located in Odessa. The limit at the Odessa North transformer is the same constraint which appears in real-time, while the limit on the Odessa North to Odessa Basin Switch 69 kV line is very similar. Three more high impact constraints from real-time also make appear on the day-ahead list. They are the Garrot to Midtown 138 kV line in Houston, the Nicole to Oak Creek 69 kV line in West Texas, and the Carrollton Northwest to Lakepoint 138 kV line in the Dallas area.

The last three constraints rounding out the list are the Twin Oak to Jack Creek 345 kV line, the Pearsall 138/69 kV Transformer and the Troys to Eddys 69 kV line. Twin Oak to Jack Creek is a constraint that limits imports to Houston from the North. The Pearsall transformer constraint was related to the high impact Valley import congestion early in the year. The Troy to Eddys constraint appeared primarily in July. Although it wasn't specifically a high impact constraint during real-time, there were other similar constraints activated in the Temple / Waco area.

In our final analysis of this section we review the most frequently occurring day-ahead constraints shown in Figure 38. This list includes the now familiar, West to North, Garrot to Midtown, Odessa transformer, and Nicole to Oak Creek constraints.



Figure 38: Most Frequent Day-Ahead Constraints

However, four constraints appearing on the list, including the top three, are constraints that would not occur in real-time. The two Laredo VFT constraints appear frequently as day-ahead

constraints, but in real-time operations all transactions with Mexico using these transformers are scheduled using a separate process which would strictly limit their volume.

The Texas Petrochemical to Texpet Switchrack 69 kV line (PR_PRS33) and the Formosa 138/69 kV transformer are both Private Use Network ("PUN") facilities. Though they are represented in ERCOT's transmission network model, because they are privately owned transmission facilities and the action necessary to relieve the constraint would be to redispatch PUN generation, they are not constraints that ERCOT should activate in real-time.

The final two constraints on the list are the China Grove Switch to Bluff Creek Switch 138 kV in west Texas and the Hickory to Locust 69 kV line in the City of Denton.

C. Congestion Rights Market

Congestion can be significant from an economic perspective, compelling the dispatch of highercost resources because power produced by lower-cost resources cannot be delivered due to transmission constraint(s). Under the nodal market design, one means by which ERCOT market participants can hedge these price differences is by acquiring Congestion Revenue Rights ("CRRs") between any two settlement points. In the zonal market only the costs associated with managing congestion on the commercially significant constraints could be hedged. All other congestion costs were uplifted to all loads.

CRRs are acquired by annual and monthly auctions while Pre-assigned Congestion Revenue Rights ("PCRRs") are allocated to certain participants based on their historical patterns of transmission usage. Parties receiving PCRRs pay only a fraction of the auction value of a CRR between the same source and sink. Both CRRs and PCRRs entitle the holder to payments corresponding to the difference in locational prices of the source and sink.

Figure 39 summarizes the revenues collected by ERCOT in each month for all CRRs, both auctioned and allocated. These revenues are distributed to loads in one of two ways. Revenues from cross zone CRRs are allocated to loads ERCOT wide. Revenues from CRRs that have their source and sink in the same geographic zone are allocated to loads within that zone. This method of revenue allocation provides a disproportionate share of CRR auction revenues to loads located in the West zone. In 2011, CRRs with both their source and sink in the West zone

accounted for 25 percent of CRR Auction revenues. This share of revenue was allocated to West zone loads, which accounted for only 7 percent of the ERCOT total.





Next, in Figure 40 we examine the value CRR owners (in aggregate) received compared to the price they paid to acquire the CRRs. Although results for individual participants and specific source/sink combinations varied, we find that in most months participants did not over pay in the auction. Across the entire year, participants spent \$370 million to procure CRRs and received \$459 million.

The two months where participants significantly overpaid to acquire CRRs were July and August. The amount paid to procure CRRs in July and August seems reasonable since it was very similar to the payments received by CRR owners in June. The smaller payment to CRR owners in July and August may have been a result of ERCOT operators choosing not to activate particular transmission constraints at times when the system was experiencing, or was anticipated to experience scarcity conditions.



Figure 40: CRR Auction Revenue and Payment Received

This may seem like an appropriate action for ERCOT operators to take so as to not reduce the amount of capacity available to the system. However, pricing parameters in ERCOT's dispatch software are set such that if there is indeed a scarcity situation, the software will allow transmission constraints to be violated if capacity is needed to meet power balance requirements. Because the dispatch software will effectively prioritize between managing transmission constraints and balancing supply and demand, we recommend that ERCOT reconsider its practice of deactivating transmission constraints in the dispatch software during peak demand conditions.



Figure 41: CRR Auction Revenue, Payments and Congestion Rent

In our next look at aggregated CRR positions, we add congestion rent to the picture. Simply put, congestion rent is the difference between the total costs that loads pay and the total revenue that generators receive. Congestion rent creates the source of funds used to make payments to CRR owners. Figure 41 presents all three values for each month of 2011. For the year, congestion rent totaled \$473 million and payments to CRR owners were \$459 million. However, in March and September through December congestion rent was less than payments to CRR owners.

We further analyze the relationship between congestion rent and payments to CRR owners by separating the impacts of CRRs that are settled based on day-ahead prices from the subset of CRRs that are paid based on real-time prices.



Figure 42: Day-Ahead and Real-Time Congestion Payments and Rent

The top portion of Figure 42 displays the comparison of day-ahead congestion rent to payments received by CRR owners. Congestion rent is larger than payments in most months, and for the year rent is \$407 million compared to \$359 million that was paid to CRR owners. The bottom portion of Figure 42 presents a different view. For this analysis we have assumed that all PTP Obligations have been fully funded from real-time congestion rent and any residual real-time congestion rent is available to fund payments to the subset of CRR owners that have elected to have their CRRs be settled based on real-time prices. With this assumption there was \$66 million in residual congestion rent available to fund real-time CRR payments of \$99 Million. Hence, real-time congestion rent is insufficient to fund all PTP Obligations and CRRs being settled in real-time. The next figure shows this explicitly.



Figure 43: Real-Time Congestion Payments

In Figure 43 the combined payments to PTP Obligation owners and CRR owners that have elected to receive real-time payments are compared to the total real-time congestion rent. For the year, real-time congestion rent was \$529 million, payments for PTP Obligations were \$463 million and payments for real-time CRRs were \$99 million, resulting in a shortfall of approximately \$33 million for the year.

This type of shortfall typically results from discrepancies between transmission topology assumptions used when clearing the day-ahead market and the actual transmission topology that exists during real-time. Specifically, if the day-ahead topology assumptions allows too many real-time congestion instruments (PTP Obligations and CRRs settled at real-time prices) with impacts on a certain path, and that path constrains at a lower limit in real-time, there will be insufficient rent generated to pay all of the congestion instruments.

From Figure 43 we can see that September and November were the months with the most noticeable deficiencies. During September there were multiple forced outages of major

transmission facilities due to wild fires. These outages could not reasonably be anticipated and resulted in significant, short duration real-time congestion. In November there was a problem with the transmission modeling around the North DC Tie that allowed PTP Obligations to be sold between two points that should have been deemed "electrically close".

For our last look at congestion we examine the impacts of the West to North constraint in more detail. Figure 44 presents the price spreads between the West Hub and North load zone as valued at three separate points in time - at the monthly CRR auction, day-ahead and in real-time.



Since CRRs are sold for one of three defined time periods, weekday on peak, weekend on peak, and off peak, Figure 44 includes a separate comparison for each.

As expected, most real-time congestion, as evidenced by the largest price spread, occurred in the off peak period, for the months of February through June and November. The day-ahead price spreads were very similar for this period, but the prices paid for CRRs were generally less than

the value received. Conversely, during the months of July through September, there was very little congestion. In July day-ahead and real-time prices were higher at the West Hub, which the results of the CRR auction did not anticipate.

IV. LOAD AND GENERATION

This section reviews and analyzes the load patterns during 2011 and the existing generating capacity available to satisfy the load and operating reserve requirements. We provide specific analysis of the large quantity of installed wind generation and conclude this section with a discussion of the daily generation commitment process.

A. ERCOT Loads in 2011

The changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. It is also important to separately evaluate the changes in the load during the highest-demand hours of the year. Significant changes in these peak demand levels have historically been very important and played a major role in assessing the need for new resources. They also affect the probability and frequency of shortage conditions (*i.e.*, conditions where firm load is served but minimum operating reserves are not maintained). The expectation of resource adequacy is based on the value of electric service to customers and the harm and inconvenience to customers that can result from interruptions to that service. Hence, both of these dimensions of load during 2011 are examined in this subsection and summarized in Figure 45.

This figure shows peak load and average load in each of the ERCOT zones from 2008 to 2011¹⁰. In each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (with about 39 percent of the total ERCOT load); the South and Houston Zones are comparable (27 percent) while the West Zone is the smallest (7 percent of the total ERCOT load).

Figure 45 also shows the annual non-coincident peak load for each zone. This is the highest load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

¹⁰ For purposes of this analysis NOIE Load Zones have been included with the proximate geographic Load Zone.



Figure 45: Annual Load Statistics by Zone

Total ERCOT load increased from 319 TWh in 2010 to 335 TWh in 2011, an increase of 5.0 percent or an average of approximately 1,800 MW every hour. Similarly, the ERCOT coincident peak hourly demand increased from 65,776 MW in 2010 to 68,379 MW, an increase of roughly 2,600 MW, or 4.0 percent.

To provide a more detailed analysis of load at the hourly level, Figure 46 compares load duration curves for each year from 2008 to 2011. A load duration curve shows the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a fairly smooth load duration curve, typical of most electricity markets, with low to moderate electricity demand in most hours, and peak demand usually occurring during the late afternoon and early evening hours of days with exceptionally high temperatures.



Figure 46: Load Duration Curve – All hours

As shown in Figure 46, the load duration curve for 2011 is significantly higher than in 2010 across all hours of the year. This is consistent with the aforementioned 5.0 percent load increase from 2010 to 2011.

To better illustrate the differences in the highest-demand periods between years, Figure 47 shows the load duration curve for the five percent of hours with the highest loads. This figure also shows that the peak load in each year is significantly greater than the load at the 95^{th} percentile of hourly load. From 2008 to 2011, the peak load value averaged 18 percent greater than the load at the 95^{th} percentile. These load characteristics imply that a substantial amount of capacity – approximately 10 GW – is needed to supply energy in less than 5 percent of the hours.



Figure 47: Load Duration Curve – Top five percent of hours

B. Generation Capacity in ERCOT

In this section we evaluate the generation mix in ERCOT. The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West Zone. The North Zone accounts for approximately 36 percent of capacity, the South Zone 28 percent, the Houston Zone 22 percent, and the West Zone 14 percent. The Houston Zone typically imports power, while the West Zone typically exports power. Excluding mothballed resources and including only 8.7 percent of wind capacity as capacity available to reliably meet peak demand, the North Zone accounts for approximately 40 percent of capacity, the South Zone 30 percent, the Houston Zone 24 percent, and the West Zone 6 percent. Figure 48 shows the installed generating capacity by type in each of the ERCOT zones¹¹.

¹¹ For purposes of this analysis, generation located in a NOIE Load Zone has been included with the proximate geographic Load Zone



Figure 48: Installed Capacity by Technology for each Zone

There were very few new units placed in service during 2011. Notable changes to ERCOT's installed generation during 2011 included two coal units being mothballed¹², additions of a new combined cycle unit and a wind unit. Even after these changes natural gas generation accounts for approximately 50 percent of the installed capacity in ERCOT.

By comparing the current mix of installed generation capacity to that in 2007, as shown in Figure 49, we can see the effects of longer term trends. Over these five years wind and coal generation are the only two categories with increased capacity. However, the sizable additions in these two categories have been more than offset by retirements of natural gas fueled steam units, resulting in less installed capacity in 2011 than there was in 2007.

¹² The mothball designation for these two units was subsequently rescinded in January 2012.



Figure 49: Installed Capacity by Type: 2007 to 2011

The shifting contribution of coal and wind generation is evident in Figure 50, which shows the percent of annual generation from each fuel type for the years 2007 through 2011. Over the five years shown, the percentage of generation produced by coal units increased slightly from 37 percent to 39 percent. Wind's generation share has increased every year, reaching 9 percent of the annual generation requirement in 2011, up from 3 percent in 2007. During the same period the percentage of generation provided by natural gas decreased from 45 percent to 40 percent.



Figure 50: Annual Generation Mix

While ERCOT has coal/lignite and nuclear plants that operate primarily as base load units, its reliance on natural gas resources drives the high correlation between real-time energy prices and the price of natural gas fuel. There is approximately 25 GW of coal and nuclear generation in ERCOT. Generally, when ERCOT load is above this level, natural gas resources will be on the margin and set the real-time energy spot price. Although coal-fired and nuclear units combined to produce more than half of the energy in ERCOT, they have historically played a much less significant role in setting spot electricity prices. However, with the recent additions of new coal generation combined with continuing increases in wind capacity, with its low marginal production, the frequency at which coal and lignite are the marginal units in ERCOT was expected to increase in the future, particularly during the off-peak hours in the spring and fall, and even more as additional transmission capacity is added that will accommodate increased levels of wind production in the West Zone. This expectation is currently tempered by the impacts on existing coal units from impending additional environmental regulations and continuing low natural gas prices.