2024 State of the Market Report for the ERCOT Electricity Markets

POTOMAC ECONOMICS

Independent Market Monitor
for ERCOT

May 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

LIST OF ACRONYMS

4CP	Four Coincident Peak	LSL	Low Sustained Limit
AORDC	Aggregate ORDC	LZ	Load Zone
AS	Ancillary Service	MIRTM	Multi-Interval Real-Time Market
ASDC	Ancillary Service Demand Curve	MISO	Midcontinent Independent System Operator
BESS	Battery Energy Storage System	MMBtu	One million British Thermal Units
CAES	Compressed Air Energy Storage	MORA	Monthly Outlook for Resource Adequacy
CARD	CRR Auction Revenue Distribution	MW	Megawatt
CC	Combined Cycle	MWh	Megawatt Hour
CDR	Capacity, Demand and Reserves Report	NCLR	Non-controllable Load Resource
CLR	Controllable Load Resource	NOIE	Non-Opt-In Entity
CMWG	Congestion Management Working Group	NPRR	Nodal Protocol Revision Request
CONE	Cost of New Entry	NREL	National Renewable Energy Lab
COP	Current Operating Plan	NSO	Notification of Suspension of Operations
CRR	Congestion Revenue Rights	NSRS	Non-Spin Reserve Service
CRRAH	CRR Account Holder	ORDC	Operating Reserve Demand Curve
CT	Combustion Turbine	PCRR	Pre-Assigned Congestion Revenue Rights
DAM	Day-Ahead Market	PFR	Primary Frequency Response
DC Tie	Direct-Current Tie	PNM	Peaker Net Margin
DME	Decision Making Entity	PPA	Power Purchase Agreement
DRRS	Dispatchable Reliability Reserve Service	PRB	Powder River Basin
DRUC	Day-Ahead Reliability Unit Commitment	PRC	Physical Responsive Capability
ECRS	ERCOT Contingency Reserve Service	PTP	Point-to-Point
EEA	Energy Emergency Alert	PTPLO	Point-to-Point Obligation with links to an Option
EIA	U.S. Energy Information Administration	PUCT	Public Utility Commission of Texas
ELCC	Effective Load Carrying Capability	PURA	Public Utility Regulatory Act
ERCOT	Electric Reliability Council of Texas	PV	Photovoltaic

ERO	Electric Reliability Organization	QSE	Qualified Scheduling Entity
ERS	Emergency Response Service	RDI	Residual Demand Index
ESR	Energy Storage Resource	RDPA	Real-Time Reliability Deployment Price Adder
EUE	Expected Unserved Energy	Reg-Down	Regulation Down Reserve Service
EV	Electric Vehicle	Reg-Up	Regulation Up Reserve Service
FFR	Fast Frequency Response	RENA	Real-Time Revenue Neutrality Allocation
FFSS	Firm Fuel Supply Service	REP	Retail Electric Provider
FFSSR	Firm-Fuel Supply Service Resource	RFI	Request For Information
FIP	Fuel Index Price	RRS	Responsive Reserve Service
GTC	Generic Transmission Constraint	RTC	Real-Time Co-optimization
GW	Gigawatt	RTCA	Real-Time Contingency Analysis
HB	House Bill	RTO	Regional Transmission Organization
HCAP	High System-Wide Offer Cap	RTOLCAP	Real-Time On-Line reserve capacity of all On-Line Resources
HE	Hour-ending	RTP	Regional Transmission Plan
HRUC	Hourly Reliability Unit Commitment	RUC	Reliability Unit Commitment
HSL	High Sustained Limit	SARA	Seasonal Assessment of Resource Adequacy
Hz	Hertz	SASM	Supplemental Ancillary Service Market
IMM	Independent Market Monitor	SB	Senate Bill
kW	Kilowatt	SCED	Security-Constrained Economic Dispatch
IRR	Intermittent Renewable Resource	SCR	System Change Request
LASCED	Look-Ahead SCED	SOC	State of Charge
LCOE	Levelized Cost of Electricity	SOM	State of the Market
LDF	Load Distribution Factor	SPP	Southwest Power Pool
LFL	Large Flexible Load	SWCAP	System-Wide Offer Cap
LFLTF	Long-Term Load Forecast	TAC	Texas Administrative Code
LMP	Locational Marginal Price	TDSP	Transmission and Distribution Service Provider
LOLE	Loss of Load Expectation	TSP	Transmission Service Provider
LOLP	Loss of Load Probability	UFR	Underfrequency Relay
LSE	Load-Serving Entity	VMP	Voluntary Mitigation Plans
		VOLL	Value of Lost Load

EXECUTIVE SUMMARY

Potomac Economics provides this State of the Market Report for 2024 to the Public Utility Commission of Texas (PUCT) in our role as the Independent Market Monitor (IMM). This report presents our assessment of the outcomes of the wholesale electricity market in the Electric Reliability Council of Texas (ERCOT). Additionally, we recommend changes to improve the competitive performance and operation of the ERCOT markets.

ERCOT manages the production and flow of electricity to more than 26 million Texas customers – about 90% of the state's total electric demand. Every five minutes, the ERCOT market coordinates the electricity output from more than 1,250 generating resources to satisfy customer demand and manages the resulting flows of power across more than 54,100 miles of transmission lines in the region. Additionally, the market prices facilitate the long-term investment and retirements of resources in the ERCOT region. Hence, the market's performance that we evaluate in this report is critical for maintaining reliability in Texas.

This report details the key trends in the ERCOT market, including the evolution of supply and demand and a review of market outcomes. Key results in 2024 include the following:

Competition and Market Power

- The ERCOT energy markets performed competitively in 2024, and the IMM found little evidence that suppliers exercised market power in the ERCOT energy market.
- Significant market power exists at both the system and zonal levels. Existing Voluntary Mitigation Plans (VMP) help mitigate the potential exercise of that market power. However, not all large suppliers currently have VMPs and the future of VMPs is not certain given their voluntary nature. Further, introduction of the Dispatchable Reliability Reserve Service (DRRS) uncertainty product in the future will create additional demand for available capacity, which may increase market power in ERCOT. Hence, a market power mitigation provision applied at the system and zonal levels may be warranted.
- In some local areas, transmission system limitations on the amount of power that can flow into the area can increase opportunities to exercise market power. However, mitigated offer price caps effectively addressed this concern in 2024.

Demand for and Supply of Electricity

- Peak demand declined by 0.3% due to less extreme temperatures in the summer, but average demand increased by 3.5%. New monthly peak demand records were set for January, April, May, October, and November.
- An influx of new supply contributed to a lower frequency of tight system conditions in 2024. The vast majority of this new capacity was solar and energy storage adding 7.5 gigawatts (GW) and 5 GW of new capacity, respectively. Most of this new capacity was built in the South and North zones.

Market Outcomes and Performance

- The ERCOT Contingency Reserve Service (ECRS) was less impactful because of increases in supply and less extreme weather conditions, but ECRS procurement and deployment practices still contributed almost $1 billion in excess real-time market costs.
- Average real-time prices excluding adders fell to $32 per MWh in 2024, a 52% drop from 2023, despite only a 14% decline in natural gas prices. Since gas units are typically marginal, energy prices (without adders) would normally track gas prices more closely. However, more than three quarters of the sharp decline in energy prices in 2024 was the result of less frequent artificial shortages caused by ECRS deployment practices.
- Price convergence was much better in 2024 compared to recent years. Prices in the day-ahead market averaged $1.46 per MWh higher than in real-time, and the average absolute difference in prices was only $17.35 per MWh, the lowest since 2020.
- Ancillary services costs dropped to $0.98 per MWh of load from $3.74 per MWh in 2023. This overall drop in the cost of ancillary services corresponds to a decrease in the average price for all products, but especially for ECRS, whose average price dropped from $76.77 per MWh in 2023 to only $9.62 per MWh in 2024. This drop was caused by an increase in supply, particularly from energy storage resources (ESRs), which contributed a substantial share of all ancillary service volume in 2024.
- Congestion costs in the real-time market totaled $1.9 billion in 2024, down 17% from 2023 and the continuation of a trend that began in 2022. Congestion cost from Generic Transmission Constraints (GTCs) increased roughly 10% from 2023 to 2024 but was still significantly lower than in 2022.

Characteristics of Supply and Demand

Changes in the demand for and supply of electricity account for many of the trends in market outcomes from year to year. We evaluate these changes and their impact on market outcomes.

Demand in 2024

Average demand for electricity in 2024 was approximately 3.5% higher than in 2023 – an increase of approximately 1,776 megawatts (MW) as the Texas economy continued to grow. Peak demand in 2024, however, was down 0.3% from 2023 because of less extreme heat in the summer. Load in West Texas continued to grow much faster than the system-wide average, up 15.7% year on year. This trend in recent years has been driven by expanding oil and natural gas production activity and thousands of megawatts of Large Flexible Loads (LFLs), large data centers used for crypto-currency mining.

Supply in 2024

In 2024, approximately 14 GW of new capacity entered commercial operation, including 7.5 GW of solar, 5.0 GW of ESRs, 1.1 GW of wind, and 500 MW of combustion turbines. Since 2020, most new capacity has come from solar and ESRs, with annual average additions of 5.2 GW and 2.4 GW, respectively. The share of generation produced from most types of generation in ERCOT were relatively flat with wind virtually unchanged and natural gas generation decreasing by less than one percent. However, changes in the resource mix and fuel prices caused the share of coal-fired generation to fall from 13.9% in 2023 to 12.6% in 2024, while solar generation rose from 7.2% in 2023 to 10.4% of all generation in 2024. In addition to solar, ESRs have also been rapidly expanding in the ERCOT market. Chapter II of this report evaluates and discusses the integration of ESRs into the ERCOT market and resulting impacts.

Review of Market Performance

ERCOT operates three distinct markets, (1) the CRR auction, wherein market participants can bid on the right to revenues derived from congestion rent, (2) the day-ahead market, which allows market participants to take financial positions on energy or to sell physical obligations to provide ancillary services, and (3) the real-time market, which sets the ultimate price of energy and manages the physical dispatch of resources while maintaining the reliability of the transmission system. We discuss the performance of each of these markets below.

Real-Time Energy Prices

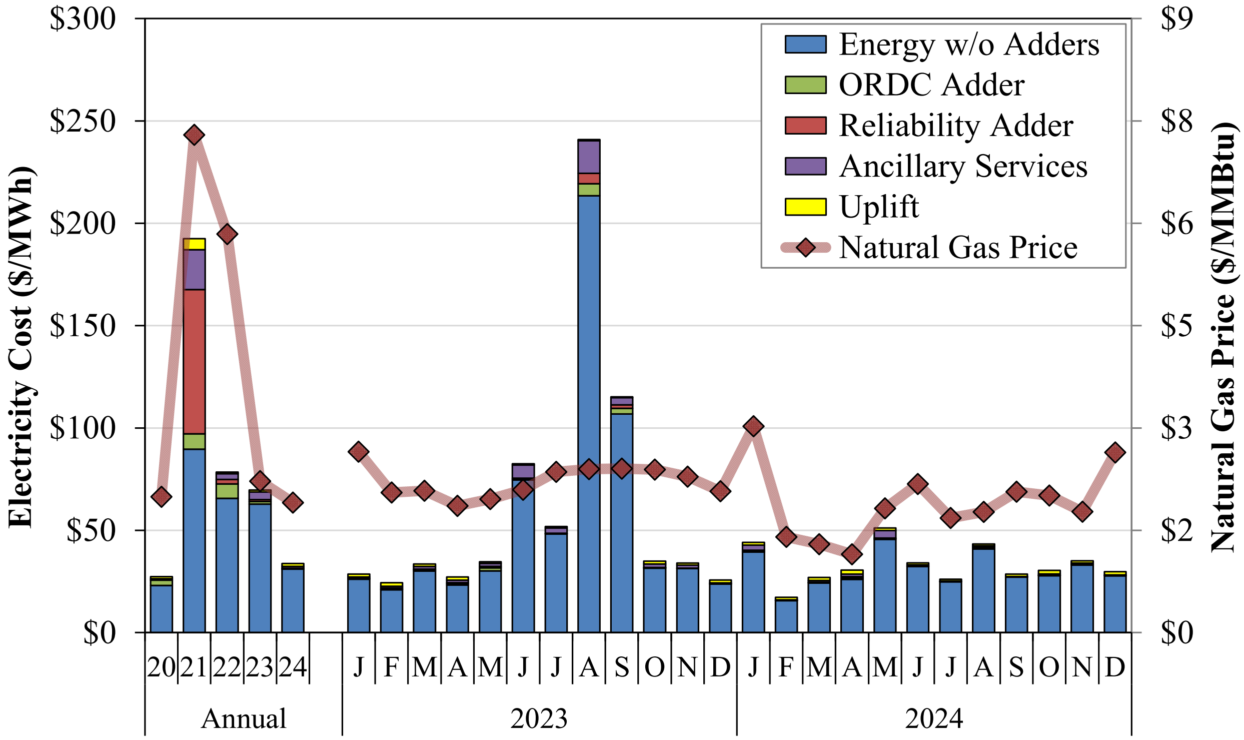
Even though only a small percentage of the energy transacted in ERCOT is settled at real-time prices, the real-time market is critical for setting expectations for future prices and informing investment decisions. Real-time prices are primarily driven by natural gas prices, because natural gas power plants are typically the marginal generator, and by shortage pricing. Shortage pricing refers to the practice of increasing market prices when operating reserves fall below certain thresholds to reflect the rising risk of insufficient supply and to signal the value of additional capacity or demand response during tight system conditions. In practice, these price adders include the Operating Reserve Demand Curve (ORDC), the Real-Time Reliability Deployment Price Adder (RDPA), and elevated marginal prices associated with ESRs or committing a quick-start generator.

Natural gas prices declined 14% from 2023 to 2024, but the price of energy excluding price adders dropped by almost 52%. The reason for this discrepancy is that there was less artificial shortage pricing caused by ERCOT's ECRS deployment practices in 2024 than in 2023. Shortage pricing was also much less significant in 2024 than in recent years. As an energy-only market, ERCOT relies heavily on high real-time prices during shortage conditions to provide economic signals for the development of new resources and retention of existing resources. ERCOT employs two different price adders to produce efficient shortage pricing, the ORDC and the RDPA.

The ORDC price adder represents the marginal reliability value of operating reserves under shortage conditions. As the level of reserves decreases, the probability of load shed increases.

That probability multiplied by the value of lost load (VOLL) should be reflected in the prices produced by the ORDC.

Average All-In Cost for Electricity in ERCOT RTM, 2020-2024



The RDPA is triggered by out of market reliability actions taken by ERCOT, such as reliability unit commitment (RUC) or deployment of demand response through the Emergency Response Service (ERS) program. The initial effect of these actions is to increase supply or decrease demand to maintain reliability, which also tends to suppress real-time prices. The RDPA is meant to reflect the pricing that would have prevailed had the reliability actions not been taken. The table below summarizes summary the price effects of the ORDC and RDPA adders.

Summary of Prices Produced by the ORDC and RDPA, 2020-2024

	ORDC			RDPA		
	Active Hours	Avg Price (Active Hours)	Avg Price (All Hours)	Active Hours	Avg Price (Active Hours)	Avg Price (All Hours)
	#	$/MWh	$/MWh	#	$/MWh	$/MWh
2020	892	$21.60	$2.61	161	$0.62	$0.01
2021	612	$91.75	$7.46	891	$565.82	$70.52
2022	1,458	$36.90	$6.94	1,664	$9.84	$2.18
2023	773	$12.12	$1.27	813	$10.23	$1.03
2024	161	$11.07	$0.25	837	$2.53	$0.24

In addition to shortage pricing produced by the ORDC and RDPA, prices above what would be expected based on the price of natural gas can also be produced by high priced marginal offers submitted by ESRs or quick-start generators. The frequency of these types of price spikes increased from 2023 to 2024, largely as a result of the increase in generation from solar resources. Generation from solar resources decreases sharply in the evening during sunset when demand for electricity is still high, particularly in the summer months. The real-time market dispatch model increasingly relies on ESRs and quick-start generation to maintain a balance of supply and demand during this solar down-ramp.

In our 2023 report, we detailed the extent to which ERCOT's implementation of ECRS in June 2023 resulted in significant excess costs caused by artificial scarcity conditions perceived by the real-time dispatch model. To recap the fundamental causes of these artificial scarcity conditions:

(1) the implementation of ECRS almost doubled the volume procured of 10-minute reserves;

(2) these reserves are withheld from the real-time energy market dispatch until manually deployed, which can cause the market to falsely perceive a shortage; and

(3) ERCOT did not effectively deploy ECRS capacity in anticipation of perceived shortages by the real-time dispatch model.

These factors resulted in frequent artificial shortage pricing when there was no shortage of operating reserves. Had these been true shortages, that would have manifested in higher prices produced by the ORDC, as was the case in 2022. This artificial shortage pricing resulted in an excess cost of more than $12 billion in 2023.

In 2024, ECRS procurement and deployment practices resulted in an excess cost of less than $1 billion. This decrease was caused by less frequent extreme temperatures in the summer and a significant increase in supply, mostly from solar and energy storage resources. Despite the decrease in excess cost from 2023 to 2024, the continuation of these ECRS deployment practices is a cause for concern. Most consumers are not directly exposed to these excess costs in the real-time market, but these pricing outcomes factor into future contracts offered by retail electric providers (REPs). Thus, we continue to recommend that ERCOT reconsider their policies for procuring and deploying ECRS, as covered in Recommendation 2023-3 of this report.

Day-Ahead Market

The day-ahead market allows market participants to take financial positions on energy in advance of the real-time market. These transactions do not result in physical obligations but allow participants to manage the risks related to real-time prices and market outcomes. Day-ahead prices averaged $31 per MWh in 2024, reflecting a modest premium over the $29 average price produced by the real-time market.

Ancillary services are operating reserves that are purchased by load serving entities in the day-ahead market. Some of these operating reserves are used to manage short-term fluctuations in supply and demand, e.g., regulation reserves, while others are procured to manage contingencies such as forced generator outages, e.g., responsive reserve service. Awards for these products amount to physical obligations to provide operating reserves in real time.

Prices for ancillary services are typically correlated with energy prices because ancillary service prices must clear high enough to make generators indifferent about providing reserves and foregoing the opportunity to sell energy. Normalized ancillary services costs dropped to $0.98 per MWh of load from $3.74 per MWh in 2023. This overall drop in the cost of ancillary services corresponds with a decrease in the average price for all products, but especially for ECRS, the average price of which dropped from $76.77 per MWh in 2023 to only $9.62 per MWh in 2024. This drop was caused by an increase in supply, particularly from ESRs, which supplied a substantial percentage of all ancillary service volume in 2024.

Transmission Congestion

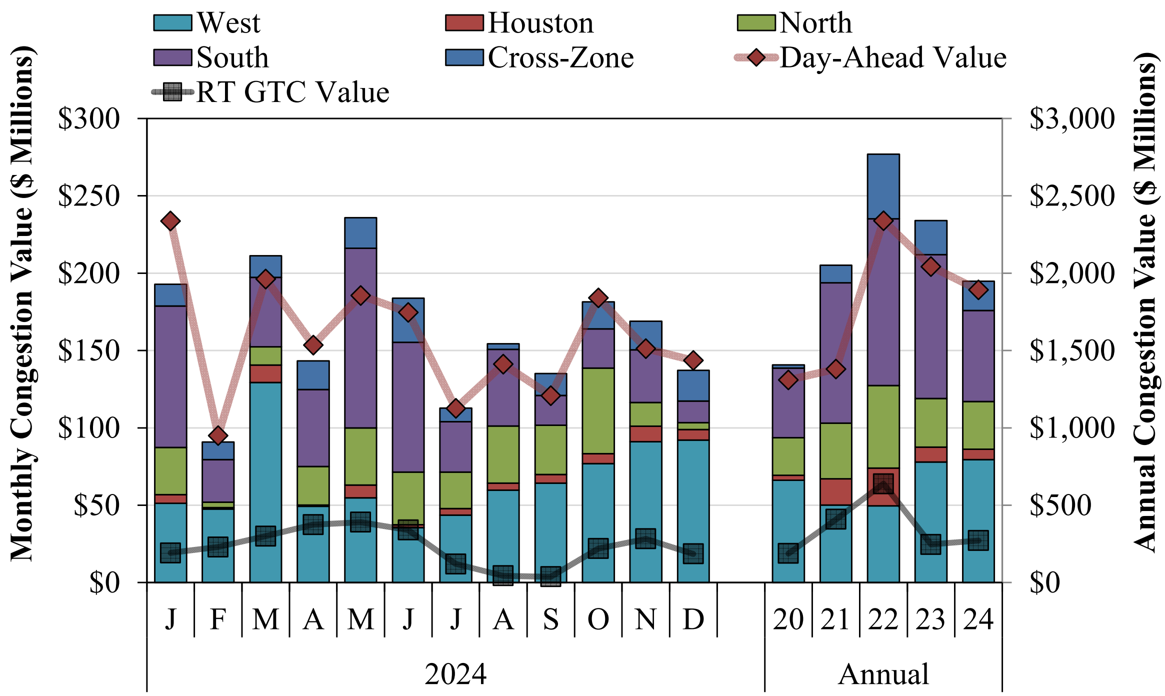
Transmission congestion arises when network power flows are restricted due to limits on transmission infrastructure such as power lines and transformers. Power flows over the network are determined by topology, i.e., the configuration of the network of transmission infrastructure, and the locations at which power is injected or withdrawn from the network. When the flow over a transmission facility reaches its limit, the market will shift generation to higher-cost units to serve load without violating the transmission constraint. Hence, congestion prevents load from being served by the lowest-cost generators. When transmission congestion occurs, the differences in costs of delivering electricity to different locations is reflected in the energy prices at each location or "node" on the network. These differences in nodal prices provide efficient economic signals for generators and consumers to produce and consume electricity at different locations. Congestion rent, which equals the difference between what is paid by consumers and the payments to generators, is based on the differences in locational prices. The financial right to this congestion rent accrues to holders of Congestion Revenue Rights (CRRs).

Real-Time Congestion Costs. To show the trends and fluctuations in congestion costs, the figure below shows real-time congestion costs by month and region for 2024 and the trend in annual costs from 2020 through 2024. The congestion costs in ERCOT's real-time market in 2024 were $1.9 billion, down 17% from 2023 and the continuation of a trend that began in 2022. The decrease in congestion can be attributed to improved congestion management, particularly through the effective use of ESRs. We provide more detail on the increasing role of ESRs in congestion management in Chapter IV. Other noteworthy congestion trends include the following:

- The West Zone surpassed the South Zone as the load zone with the highest intra-zonal congestion costs. The West Zone was also the only load zone that experienced an increase in intra-zonal congestion costs from 2023 to 2024.

- Several new transmission facilities were energized in the South Zone in 2024 that contributed to the decrease in intra-zonal congestion compared to 2023.
- Total congestion from GTCs increased by roughly 10% from 2023 to 2024 but was still significantly lower than in 2022.
- Real-time congestion costs were 3% higher than in the day-ahead market, largely the result of higher real-time congestion costs associated with GTCs, which are often the result of forecast error for generation from renewables.

Real-Time Transmission Congestion Costs, 2020-2024



Congestion Revenue Rights. Participants can hedge congestion costs in the day-ahead market by purchasing CRRs. CRRs are economic property rights that entitle the holder to the day-ahead congestion rent between two locations on the network. They are sold in auctions administered by ERCOT in monthly blocks as much as three years in advance. The revenues collected through the CRR auction are distributed back to load-serving entities to reduce the costs of paying for the transmission system. CRR auction revenues continued to rise in 2024, increasing to $1.7 billion in 2024 compared to $1.4 billion in 2023.

Despite the continued increase in revenues from CRR auctions, payments to CRR holders decreased for the second year, making payments to CRR holders, which were just under $1.7 billion in 2024, less than CRR auction revenue for the first time since 2015. This convergence between auction revenues and payments to CRR holders indicates that the value of CRRs is increasingly reflected in the bidding and clearing prices of the auction. There were no shortfalls in CRR payments except for $23 million in derated target payments as described in Chapter IVIV.

Competition and Market Power

We evaluate market power from two perspectives: structural (does market power exist?) and behavioral (have attempts been made to exercise it?). Based on our analysis, we find that significant structural market power continued to exist in 2024, but there is little evidence that suppliers abused market power in the real-time energy market. We identified a concern with non-competitive outcomes in the Non-Spin Reserve Service (NSRS) product in the 2022 Report and changes to the VMPs of larger suppliers were made in response to this concern.

Structural Market Power

For electricity markets, a more effective indicator of potential market power than traditional market concentration metrics is to analyze when a supplier is "pivotal." A supplier is pivotal when its resources are needed to satisfy customer demand or reduce flows over a transmission line to manage congestion. The results below indicate that market power continues to exist in ERCOT and requires mitigation measures to address it. Over the entire ERCOT region:

- At least one pivotal supplier existed at the system-level in 63% of all hours in 2024, compared to 50% in 2023 and 57% in 2022.
- Under high-load conditions, a supplier was pivotal in roughly 90% of the hours.
- Market power at the zonal level is also a concern, with frequency of at least one pivotal supplier within a zone as high as 66% in 2024.

The frequency of conditions exhibiting structural market power raises concern. The VMPs that are in place do provide some protection against the exercise of market power. However, not all larger suppliers have adopted a VMP. We provide a recommendation to pursue market power mitigation at the system and zonal level later in this section.

Market power can also be a much greater concern in local areas when power flows over the network cause transmission congestion that isolates these areas. Market rules cap prices that suppliers can offer in these cases, mitigating suppliers' ability to exercise local market power.

Behavioral Evaluation

We also evaluate behavior to assess whether suppliers engaged in withholding to increase prices. Economic withholding occurs when a supplier raises its offer prices to levels well above the expected marginal cost to produce electricity. Physical withholding occurs when a supplier makes a resource unavailable. Either of these strategies will reduce output from the withheld resource and thereby increase the prices paid to the supplier's other resources.

- *Economic withholding*. Our output gap metric used to measure potential economic withholding – the quantity of economic energy that is not produced by online resources – showed moderate quantities of potential economic withholding in 2024.
- *Physical withholding*. Both large and small suppliers made more capacity available on average during periods of high demand in 2024 by minimizing planned outages and maximizing the generation offered from each resource. These results together with our ongoing monitoring indicate few potential physical withholding concerns.

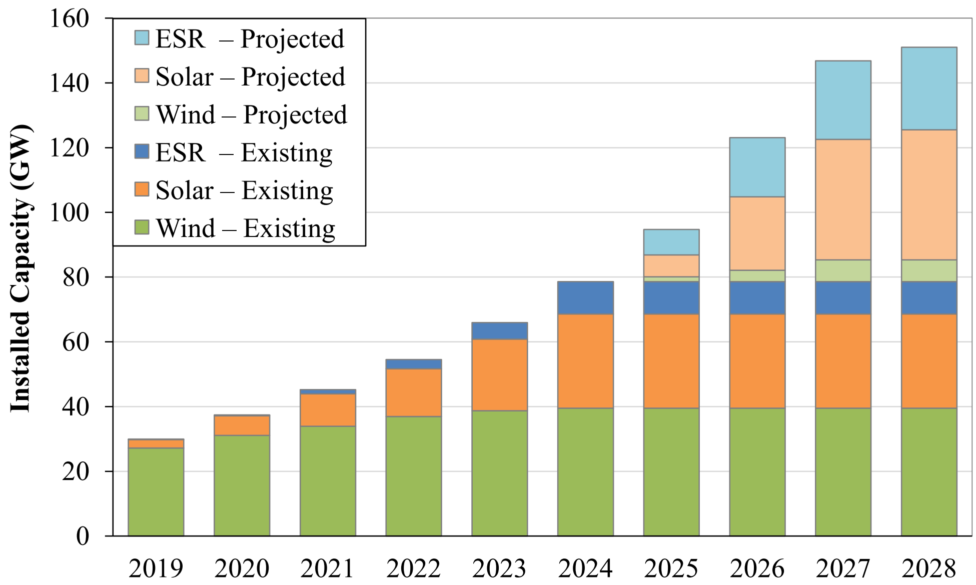
Self-commitments by a large supplier continued to lag previous self-commitment levels, which is likely due to incentives caused by ERCOT's use of RUC. Two market rule revisions have been implemented that reduce such incentives.[1]

Evolution of Supply and Demand in ERCOT

Changes to ERCOT's Supply Portfolio

Solar and ESRs comprised the largest share of new capacity additions in ERCOT in 2024. This trend is expected to continue over the next several years, as shown in the following figure. By the end of 2024, more than 29 GW of solar and 10 GW of ESRs had been installed. An additional 40 GW of solar and 25 GW of ESRs have already signed interconnection agreements for installation by 2028.

Development of Renewable Resources and Energy Storage, 2019-2028



1 Nodal Protocol Revision Request (NPRR) 1092, Reduce RUC Offer Floor and Limit RUC Opt-Out Provision was filed by the IMM and approved by the Board. The RUC offer floor was reduced to $250 per MWh but the RUC opt-out provision will be removed once ERCOT completes implementation. NPRR 1172, Fuel Adder Definition, Mitigated Offer Caps, and RUC Clawback, was implemented by the PUCT on March 1, 2024.

As the resource mix in ERCOT continues to shift, this transformation will carry several implications for market performance and reliability:

- **Shift of Peak Net Load into Evening Hours:** Peak net load refers to the peak of system demand net of renewable generation. As more solar is added to the grid, these hours are increasingly shifting later in the day. By 2028, the peak net load hour is expected to occur in the late evening, when solar output is minimal or zero, but demand is still high.
- **Changing Capacity Factors During Peak Net Load:** Historically, solar resources have aligned well with peak demand, particularly in the afternoon. As the net load peak moves into the evening, solar's contribution during these evening hours falls to zero. Meanwhile, wind resources, which typically begin ramping up as the sun sets, have higher capacity factors during these increasingly critical hours. This hand-off between solar and wind can result in reliability risk if wind output lags its forecast.
- **Growing Absolute Risk from Expanding Solar Output:** Solar resources are being developed across a broad geographic footprint, which reduces the risk that localized cloud cover will significantly impact total solar output. However, the rapid growth of solar as a share of the overall generation mix means that forecasting errors now have greater absolute consequences for system reliability.

Indeed, forecast errors for intermittent renewables directly impact thermal unit commitment. Under-forecasting generation from renewables may result in inefficient economic outcomes, but over-forecasting can lead to reliability issues if thermal units are not committed in time to compensate. This dynamic is particularly challenging during hot summer evenings, when solar output declines rapidly and demand remains elevated.

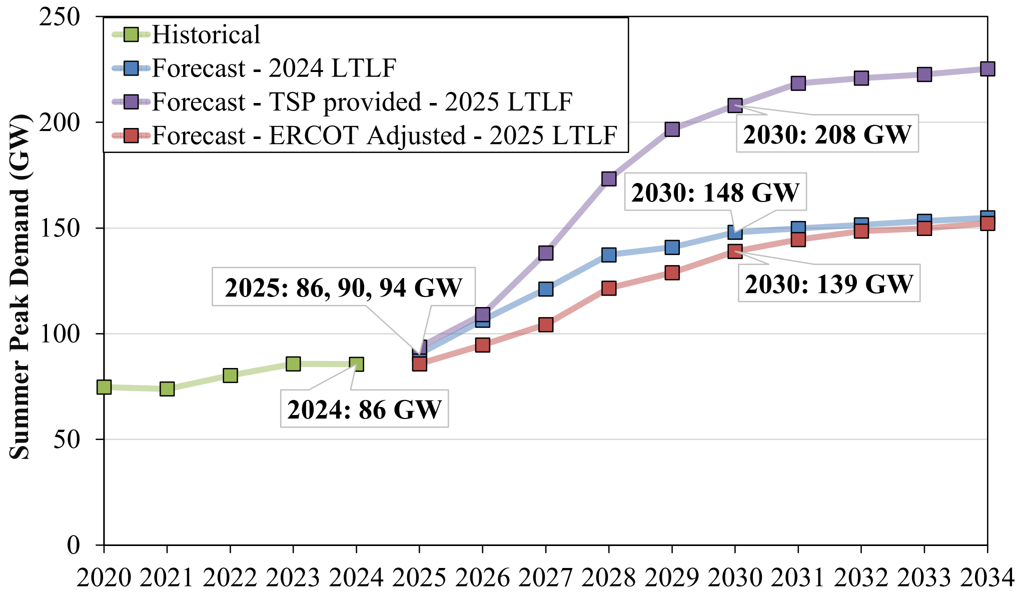
ESRs are well-suited to manage this kind of reliability risk. They can quickly ramp up output to meet demand during brief shortfalls or bridge the evening transition between declining solar and ramping wind. However, ERCOT's current dispatch engine, Security-Constrained Economic Dispatch (SCED), optimizes ESR state of charge one interval at a time (i.e., five minutes ahead), limiting its ability to coordinate ESR dispatch for longer-duration needs. Implementing a Multi-Interval Real-Time Market (MIRTM), as proposed in Recommendation 2022-1, and setting appropriate ancillary service duration requirements for ESRs, as outlined in Recommendation 2024-2, would significantly enhance the system's ability to use ESRs effectively.

While ESRs provide important reliability benefits during summer peak periods, they remain less effective for managing reliability risks associated with extreme winter conditions. For prolonged events like Winter Storm Uri, ESRs may have limited usefulness if they deplete their state of charge early into the event. Thus, these resources are not well-suited for managing multi-day cold-weather reliability events. Addressing these limitations will require further improvements to market design, including the development of an uncertainty product, as outlined in Recommendation 2021-2.

Projected Load Profile in ERCOT

ERCOT has experienced significant load growth in the past five years and is projected to have an amplified load growth over the coming five years. Projected peak load in 2030 is 148 GW according to ERCOT's 2024 load forecast, which is a 72% increase over the 2024 peak load of 86 GW. This annualized growth rate of roughly 10% per year for six years far exceeds the normal load growth between 1% and 3% per year that Regional Transmission Organizations (RTOs) typically experience. ERCOT's 2025 load forecast projects an even higher summer peak demand of 208 GW by 2030, although ERCOT adjusted this forecast to 139 GW, which still implies an average load growth rate of 8.5% annually for the next five years.

ERCOT Long-Term Peak Demand Forecast



Accurately projecting load growth five to ten years into the future is inherently difficult. In ERCOT, this challenge is heightened by two factors. First, the rapid influx of data centers is a relatively recent development, leaving little historical data to inform forecasts. In addition, decisions to build new data centers are highly sensitive to changing economic conditions and competition from other regions seeking to attract similar investments. Second, the 88th Texas Legislature, through House Bill (HB) 5066, mandated the inclusion of loads for which an electric utility has yet to sign an interconnection agreement, which resulted in a data collection process for anticipated load growth that likely overstates what can reasonably be expected. Recognizing this, ERCOT has included multiple scenarios in its Capacity, Demand, and Reserves (CDR) Report to illustrate a range of plausible load growth outcomes in the coming years.

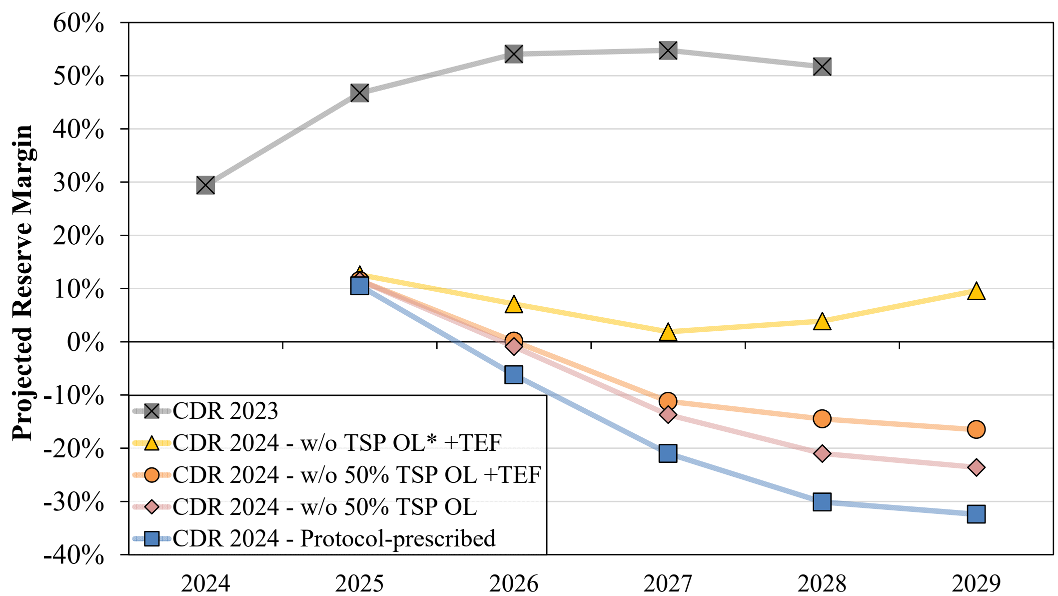
Data centers fall into two general categories that differ in how they interact with the electricity market. Cryptocurrency mining operations are typically flexible and highly sensitive to electricity prices. These facilities can curtail usage when prices are high and ramp up when prices are low, which makes them valuable as price-responsive load. In contrast, cloud computing data centers, including those supporting artificial intelligence workloads, are far less flexible. They operate continuously and are largely insensitive to electricity prices, which limits their ability to provide demand response. Nearly 80% of the increase from the 2024 load forecast to the 2025 forecast, or approximately 49 GW, came from cloud-computing data center growth alone.

While much of the recent load growth has come from cryptocurrency mining, future data center development is expected to be dominated by less flexible cloud computing facilities. Moreover, existing cryptocurrency mining enterprises may convert to conventional data center operations if cryptocurrency prices decline or electricity prices rise. This shift would reduce the system's demand-side flexibility and increase the need for generation to maintain reliability. These trends and their implications for resource adequacy are discussed in more detail in Section VI.

Planning Reserve Margin

The uncertain economic outlook for various types of load has led to a wide range of projections for resource adequacy in ERCOT. One of the primary ways ERCOT communicates these projections is through the planning reserve margin. This metric measures the difference between total available generation capacity and expected peak demand, expressed as a percentage of peak demand. It serves as a basic indicator of whether the system is expected to have sufficient capacity to meet demand under typical conditions. Higher reserve margins suggest more cushion in the system, while lower margins point to tighter supply and elevated risk of shortages. The following figure compares the reserve margin projections from the December 2023 and December 2024 CDR reports. The gray series at the top represents the December 2023 CDR report while the remaining data series at the bottom represent different scenarios from the December 2024 CDR report.

Planning Reserve Margin, CDR 2023 vs CDR 2024



*OL = Officer Letter

These two reports present sharply different expectations for resource adequacy in ERCOT over the next five years. While some of the differences are the result of procedural changes in how the CDR is compiled, as discussed further in Chapter VI, the steep drop in reserve margin projections in the 2024 CDR highlights two major limitations of the CDR itself: (1) the over-projection of future load and (2) the under-projection of future capacity. The May 2025 projects even tighter conditions for 2030, projecting reserve margins between -20% and -50%.

The over-projection of load stems primarily from ERCOT's interpretation of HB 5066, which mandates the inclusion of officer letter loads into ERCOT's load forecast; ERCOT implemented this directive by including all potential large loads regardless of the likelihood that this load will materialize. Until recently, ERCOT's methodology did not apply any attrition factor to account for load projects that may be speculative or ultimately canceled. At the same time, the CDR under-projects capacity by only including generation resources that have already signed interconnection agreements. This excludes many projects that are likely to contribute to resource adequacy within the planning horizon but have not yet reached that administrative milestone. As a result, the CDR functions more as an accounting ledger of formally documented metrics than a predictive model of future system conditions. Again, these limitations, and their implications for reliability forecasting, are discussed in greater detail in Chapter VI.

Future Market Design Improvements

In light of the evolution of supply and demand in ERCOT, several market design improvements are necessary to send efficient market signals to promote reliability. We discuss several future market design improvements in the following subsections.

Shortage Pricing and Procurement Methodology for AS under RTC

In the current market design, ancillary services are procured in the day-ahead market according to a fixed Ancillary Service (AS) Plan that sets reserve quantities for each hour. The day-ahead market always procures the full volume of reserves defined by the AS Plan, and the real-time market treats these reserves as inaccessible unless they are manually released to the dispatch model. Shortage pricing is then determined by the ORDC outside of the dispatch model based on the level of available operating reserves.

Real-time co-optimization (RTC), set to be implemented in December 2025, will improve many of the issues we raised regarding market performance and operational risks. RTC allows the real-time market to simultaneously schedule energy and operating reserves. A critical aspect of the logic of RTC is that the market is allowed to go short on real-time operating reserves, and the cost of this shortage is defined by a set Ancillary Service Demand Curves (ASDCs). This feature of RTC enables it to more efficiently manage uncertainty than the current market design where reserves are inaccessible until they are manually deployed.

Consider the solar down-ramp scenario described in the previous section. When solar production falls short of forecast, the system experiences a drop in operating reserves. In the current design, this shortfall would typically prompt a manual reserve deployment. Under RTC, however, the drop in reserves leads the real-time market to prioritize energy dispatch over holding reserves. This results in a reserve shortfall that is priced according to the ASDCs. In this way, allowing the market to go short on reserves replaces the manual practice of releasing reserves for energy dispatch. Embedding this trade-off in the real-time market clearing logic will address many of the issues we have identified with ECRS deployment, as discussed in Recommendation 2023-3.

We propose additional recommendations for ERCOT's shortage pricing mechanism and ancillary service methodologies, as outlined in Recommendations 2024-1a and 2024-1b. In the first, we recommend that the ASDCs be reformulated based on the marginal reliability value of each product. In the second, we recommend that ERCOT incorporate a stochastic risk methodology for setting target levels for operating reserves. These two recommendations in concert would result in a set of ASDCs that accurately reflects the reliability value of operating reserves, resulting in efficient shortage pricing and promoting resource adequacy.

Uncertainty Reserve Product

The significant increase in the penetration of intermittent renewable resources has introduced operational reliability issues stemming from uncertainty in the forecast for renewable generation. Over-forecasted generation from renewable resources can result in under-commitment from thermal generators, which can lead to tight system conditions. ERCOT has responded to these uncertainties by increasing its use of the RUC process and procuring more 30-minute reserves. Both of these strategies distort market outcomes and result in excess costs to consumers in the form of uplift. Therefore, we have recommended that ERCOT implement a longer-term reserve product to be provided by resources that can start in two hours or less when uncertainties manifest that may threaten reliability. For more detail, see Recommendation 2021-2.

Multi-Interval Real-Time Market

One of the primary purposes of the real-time market is to efficiently dispatch online resources to serve forecasted load. A shortcoming of the ERCOT real-time market process is that it only considers a single interval five minutes into the future for determining optimal dispatch instructions. Implementing a real-time market process that can look ahead and optimize across several intervals, i.e., a MIRTM has several benefits that are particularly valuable given the rapid growth of intermittent renewable resources (IRRs) and ESRs. For more detail, see Recommendation 2022-1.

Transmission Cost Allocation

The cost of expanding and maintaining the transmission system in ERCOT is allocated to consumers using the four coincident peak (4CP) method, which calculates each consumer's transmission tariff rate for the following year based on their load ratio share during the 15-minute intervals with the highest system-wide demand in the prior year from June through September.[2] This methodology was originally designed to allocate costs to those driving system-wide summer peaks under the assumption that peak demand was the primary driver of transmission investment. However, the drivers of transmission investment have changed. Today, new transmission is increasingly built to connect areas of load growth with geographically distant sources of generation, particularly renewable resources. Contribution to system-wide peak demand no longer reflects how or where transmission costs are incurred. For this reason, we continue to recommend that policymakers move away from 4CP and implement a transmission cost allocation framework that more accurately reflects cost causation. For more information, see Recommendation 2015-1.

System-Wide and Zonal Market Power Mitigation

There is a high degree of structural market power that exists in the ERCOT system as well as at the zonal level. This provides a significant opportunity for suppliers to exercise market power throughout the year and raise prices above competitive levels. The VMP mechanism can be an effective means to address this issue. However, not all suppliers with market power have committed to a VMP and the terms of the VMPs must reflect the expectation of a competitive offer. Given the extent of market power in the ERCOT system, we advise that a three-step real-time market power mitigation system be put in place for both system-level and zonal-level market power. The proposed system would be run in-line with the real-time market. It should assess structural market power, then evaluate whether portfolios that possess market power are offering energy or ancillary services at prices above a competitive level. Finally, it should determine whether the uncompetitive offers would have a material impact on market prices. A mitigated offer price would be substituted for the original offer in for portfolios where all three of these tests returned the affirmative.

Resource Adequacy through Markets

The extremely high projected load growth, coupled with reliability issues observed during recent winter extreme weather conditions, has focused the attention of policy makers on how to achieve future resource adequacy needs. There are limited general market approaches that can be taken to provide incentives for new generation to meet future needs.

- The current energy-only market in ERCOT relies in large part on expected shortage revenues to drive investment in new resources, which can be calibrated to strengthen investment incentives and support resource adequacy.
- A centralized capacity market to directly procure the capacity needed to satisfy resource adequacy targets has been utilized by a number of RTOs in the eastern interconnection.

[2] 16 Tex. Admin. Code §25.192. Transmission Service Rates: http://www.puc.texas.gov/agency/rulesnlaws/subrules/electric/25.192/25.192.pdf

- A decentralized load obligation model that requires load-serving entities (LSEs) to self-supply or bilaterally procure sufficient capacity to meet resource adequacy requirements.

All market-based proposals would fall within one of these three general approaches. The report discusses the pros and cons of each approach, as well as the details that must be considered in developing a market-based solution to support resource adequacy. Recent efforts to identify appropriate market constructs to achieve resource adequacy have revealed a tension between allowing shortage conditions and associated revenues to rise versus approaches to prevent shortage pricing. Given projected load growth and the period of development between price signals and commercial operation of new generation, ERCOT will need to prioritize development of market solutions for ensuring resource adequacy.

Recommendations

Each year, we produce a set of recommendations aimed at improving market efficiency, enhancing reliability, and mitigating the potential for market participants to exercise market power. Some recommendations are new and reflect emerging challenges or developments in the current year. Others carry over from prior years because they remain relevant and unaddressed. In some cases, past recommendations are retired because they have been successfully implemented, are no longer applicable, or have been folded into a broader or updated recommendation. Each recommendation is numbered based on the year it was first introduced, followed by the order in which it is presented in this report.

Number	Recommendation Title
New Recommendations to Improve Market Performance	
2024-1	Reform Shortage Pricing Mechanism and AS Methodology to Reflect Reliability Risk
2024-1a	Define ASDCs According to Marginal Reliability Value of Each Product
2024-1b	Adopt a Probabilistic/Stochastic Risk Methodology for the AS Plan
2024-2	Set Duration Requirements for ECRS and Non-Spin Reserve Service (NSRS) to One Hour
2024-3	Implement Process to Mitigate Market Power at System and Zonal Levels
2024-4	Establish Real-Time Offer Requirements, Penalties, and Proxy Pricing
Recommended Market Improvements from Prior Years	
2023-3	Improve the Procurement and Deployment of ECRS
2023-4	Align FFSS Pricing and Deployment Practices with Market Operations
2022-1	Implement a Multi-Interval Real-Time Market
2021-1	Eliminate the “Small Fish” Rule
2021-2	Implement an Uncertainty Product
2020-3	Reconfigure Load Zones to Reflect Prevailing Congestion Patterns
2020-4	Implement a Point-to-Point Obligation Bid Fee

2019-2	Price Ancillary Services Based on the Shadow Price of Procuring Each Service
2015-1	Modify the allocation of transmission costs by transitioning away from the 4CP method.
Recommendations Being Retired	
2023-1	Increase a Constraint's Shadow Price Cap in Real-Time When Appropriate
2023-2	Revise Proxy Offers for IRRs Without Energy Bids
2022-3	Allow Transmission Reconfigurations for Economic Benefits
2022-4	Change the Linear Ramp Period for ERS Summer Deployments to 3 Hours
2022-5	Change Historical Lookback Period for the ORDC Mean and Standard Deviation Calculations
2021-3	Reevaluate net metering at certain sites

New Recommendations to Improve Market Performance

2024-1 Reform Shortage Pricing Mechanism and AS Methodology to Reflect Reliability Risk

For the shortage pricing mechanism to function properly, the ASDCs should be designed to reflect the marginal reliability value that each reserve product contributes to the system. A well-structured ASDC framework incorporates the following elements:

1. The reliability value of operating reserves should be determined using a probabilistic assessment of the specific risks that each reserve product is intended to manage, such as forecast errors, forced outages, or ramping needs.

2. Each ASDC should slope downward until the probabilistic methodology indicates that additional reserves provide no further reliability benefit. Beyond this point, reserves should be valued at zero, signaling that they offer no marginal contribution to system security.

3. The marginal reliability value used to shape the ASDCs should be reviewed and updated on a regular basis to ensure it reflects changing system conditions, evolving resource capabilities, and updated risk assessments.

We split this recommendation into two parts that both address a fundamental aspect of RTC, i.e., the appropriate valuation of real-time operating reserves. The first (2024-1a) recommends that the ASDCs be designed based on the marginal reliability value of each product. The second (2024-1b) recommends implementing a probabilistic risk-based methodology for setting the target volume of real-time operating reserves. These recommendations should be implemented in tandem to ensure consistency in how operating reserves are valued across both pricing and procurement mechanisms.

2024-1a Define ASDCs According to Marginal Reliability Value of Each Product

RTC is currently scheduled to go live in ERCOT in December 2025. This transition will replace the prevailing ORDC with a set of ASDCs and implement real-time co-optimization of energy and ancillary services. Unlike today's approach, where shortage pricing is added after the dispatch solution, RTC will embed shortage pricing into the optimization itself based on the formulation of the ASDCs. Unfortunately, the current formulation of ASDCs does not link the shortage price for an ancillary service to its value in reducing system-wide reliability risk, weakening the signal that shortage pricing is meant to send.

Over time, the prevailing ORDC has been modified to meet goals such as ensuring revenue sufficiency and limiting consumer exposure to extreme costs. These modifications have shifted shortage pricing away from its intended function, which is to reflect the marginal reliability benefit of operating reserves. In the RTC framework, that disconnect becomes even greater because of the decision that ASDCs should aggregate into a curve called the Aggregate ORDC (AORDC) that resembles the prevailing ORDC. This constraint prevents each ancillary service from being valued based on its specific contribution to reliability. We recommend removing the requirement that the ASDCs aggregate to the AORDC and that each ASDC be constructed individually to reflect the reliability value provided by the corresponding reserve product.

2024-1b Adopt a Probabilistic/Stochastic Risk Methodology for the AS Plan

Since 2022, ERCOT has procured significantly more operating reserves than other ISO or RTO. This over-procurement stems largely from an AS Methodology that fails to reflect the probabilistic risks operating reserves are intended to manage. As part of the AS Study mandated by PURA 35.004(g), we demonstrated how a probabilistic, stochastic risk-based approach could be used to set more appropriate targets for real-time operating reserves. This methodology contrasts with ERCOT's current deterministic method, which simultaneously results in excess reserves for most hours of the year but fails to account for the risks posed by forced outages or extremely volatile weather conditions.

Our approach used historical supply and demand data to simulate the probability of load shed at various levels of real-time operating reserves. It captured both net load forecast errors, which can lead to inefficient commitments, and forced outages, which immediately reduce supply. It also incorporates a time horizon for forecast error that is consistent with the short-term reliability risks that operating reserves are meant to manage.

Our analysis found that ECRS and NSRS procurement volumes could be substantially reduced while maintaining a load shed probability of just 5% per year. Moreover, average reserve volumes could be further reduced if part of the AS Plan were set dynamically and closer to real time. Today's practice of setting targets on a month-hour basis for the entire year inflates reserve needs in most hours to cover a small number of high-risk intervals. Historically, consumers have opposed dynamic reserve planning because it limits hedging opportunities, but the savings from lower reserve volumes could justify a transition toward a more flexible procurement schedule.

We acknowledge that the methodology we presented does not capture every reliability risk operating reserves are meant to address. It does not account for transmission-related issues or faster contingency events that fall outside the scope of ECRS and NSRS. Additionally, while our model relies on historical system conditions, any practical implementation would need to incorporate forward-looking adjustments to reflect changes in load and resource mix. We recommend that ERCOT refine this approach to address these limitations.

2024-2 Set Duration Requirements for ECRS and NSRS to One Hour

To provide ECRS or NSRS, ERCOT currently requires ESRs to maintain two or four hours of state of charge, respectively, based on the size of their obligation. For example, a 10 MW award for NSRS would require 40 MWh of available energy. We recommend reducing these duration requirements to one hour for both products. This change is justified for two reasons. First, a one-hour requirement better aligns with the short-term reliability risks that ECRS and NSRS are designed to address. Second, longer duration requirements create incentives for ESR offer behavior that can undermine reliability during shortage conditions.

Longer duration requirements for ESRs tend to distort incentives during scarcity conditions. Requiring two to four hours of state of charge for ESRs to provide operating reserves effectively reduces the extent to which ESRs can provide reserves and increases the likelihood of shortage pricing for energy and ancillary services. Higher energy prices, in turn, push ESRs to shift from providing reserves to delivering energy earlier in a reliability event, since the opportunity cost of holding reserves is greater than the value of selling energy. For instance, if energy prices are peaking, a two-hour reserve obligation implies an opportunity cost twice the energy price. Conversely, the higher reserve prices will tend to incentivize thermal power plants to provide reserves rather than energy, because their cost for providing reserves is lower than their cost of producing energy. This dynamic runs counter to promoting reliability, as it would be preferable for gas turbines to be dispatched for energy before duration-limited resources. Ideally, duration-limited ESRs should be prioritized for reserve provision, while thermal units should be dispatched for energy.

In April 2025, ERCOT submitted NPRR 1282,[3] which would reduce the duration requirement for ECRS to only one hour but would keep the duration requirement for NSRS at four hours. This proposal represents a step in the right direction, but we maintain that the duration requirement for NSRS should also be reduced to one hour. We recognize the need for some longer-duration operating reserve capacity to manage uncertainty of supply and demand under extreme or volatile weather conditions. That is why we continue to recommend an uncertainty

3 https://www.ercot.com/mktrules/issues/NPRR1282

product as described in recommendation 2021-2, and we think DRRS could more effectively serve this purpose than the current practice of imposing duration requirements longer than an hour onto ECRS and NSRS. In summary, we recommend that the duration requirements for ECRS and NSRS be set at one hour and that ERCOT implement DRRS as soon as practicable to serve as a longer duration uncertainty product.

2024-3 Implement Process to Mitigate Market Power at System and Zonal Levels

Given the high frequency of market power at the system and zonal levels in ERCOT and gaps that exist with the existing VMP mechanism, we recommend application of a three-step market power mitigation framework to ensure competitive market outcomes. The three-step approach should assess whether (i) a market participant possesses market power (structural test), (ii) a market participant with market power has attempted to exercise market power via economic withholding (behavioral test), and (iii) the attempt to exercise market power would have a material effect on price (impact test). For the group of market participants that fail these three tests, a competitive offer should be substituted for the original offer in the running of the real-time market.

2024-4 Establish Real-Time Offer Requirements, Penalties, and Proxy Pricing

QSEs are responsible for submitting offer curves that specify the prices and quantities at which their resources are willing to provide energy and, with the implementation of RTC, ancillary services in the real-time market. Under the current market design, if a QSE submits an incomplete energy offer curve, ERCOT substitutes a proxy offer to maintain a solvable dispatch solution. These proxy offers are set at $1,500 for IRRs and at the system-wide offer cap (SWCAP) for non-IRRs, creating steep, non-competitive offer curves that lead to inefficient and costly dispatch outcomes.

With the upcoming implementation of RTC, the potential impacts of incomplete offers could become more significant. Whereas today, ancillary services are only procured in the day-ahead market, under RTC, ancillary services and energy will be co-optimized every 5 minutes, making complete real-time offers for AS critical for securing reserves. Incomplete AS offers limit access to available reserves and result in steep proxy offers that inflate scarcity pricing and distort dispatch. Currently, there is no penalty for submitting incomplete offers. As a result, QSEs may either intentionally withhold offers or mistakenly fail to submit complete offers.

In response to these concerns, NPRR 1269 proposed that any incomplete AS offer should be replaced with a proxy offer equal to the lesser of $2,000 or the ASDC price at the 95% procurement level. We believe this methodology risks setting AS clearing prices significantly above competitive levels, resulting in unnecessary market costs. The PUCT approved NPRR 1269 in May 2025.

To address this issue, we recommend the following:

1. Proxy offers should reflect competitive pricing for energy and ancillary service products.
2. ERCOT should make its must-offer requirement explicit for all QSEs.
3. Penalties should be imposed on QSEs that fail to submit complete offers.
4. Offers constructed by proxy should be flagged in SCED data to allow continuous evaluation of their effect on market performance.

Recommended Market Improvements from Prior Years

2023-3 Improve the Procurement and Deployment of ECRS

When ECRS was implemented in June 2023, ERCOT did not establish clear and consistent deployment criteria. This lack of transparency contributed to artificial shortages, as operators often delayed releasing ECRS into the real-time market until after scarcity pricing had already begun. As a result, ECRS deployment practices caused significant unnecessary costs to the market during 2023.

In 2024, the market impact of ECRS inefficiencies was smaller than in the prior year. Two primary factors contributed to this improvement. First, milder summer conditions reduced the frequency and severity of grid stress events. Second, increased supply from new solar and energy storage resources added flexibility to the system, allowing ERCOT to manage real-time conditions with fewer shortages. Together, these factors resulted in excess ECRS-related costs of less than $1 billion in 2024, compared to the $12 billion estimated for 2023. Nevertheless, deployment practices remain a source of inefficiency that warrants further attention.

In March 2024, ERCOT proposed NPRR 1224 to address concerns with ECRS deployment. NPRR 1224 would have established formal deployment criteria based on real-time market conditions, deploying in 500 MW increments when power balance violations exceeded 40 MW for 10 consecutive minutes. Additionally, the capacity released to SCED would be subject to a $750 per MWh energy offer floor, for which there is no reasonable basis. Hence, we found this proposal to be flawed and it was ultimately rejected by the Commission in July 2024. However, ERCOT still implemented the deployment process proposed by NPRR 1224, but without the $750 per MWh energy offer floor.

Once RTC is implemented in December 2025, these deployment issues will largely be resolved. Under RTC, the real-time market will automatically go short on reserves according to the prices set by the ASDCs, effectively eliminating the need for manual ECRS deployment decisions. Moreover, several of our other recommendations regarding ancillary services under RTC are intended to further ensure that reserve procurement and deployment align with real-time reliability risks and system conditions.

Until RTC is implemented, however, the current ECRS deployment criteria still present a risk to efficient market outcomes. Artificial shortages and unnecessary scarcity pricing can still occur if ECRS is not released early enough to meet forecasted conditions. We recommend that ERCOT further refine its deployment practices by using the existing look-ahead SCED (LASCED) tool. Using LASCED outputs, ERCOT could deploy ECRS proactively based on forecasted power balance violations, rather than waiting for real-time shortages to materialize. Implementing such a forward-looking deployment trigger would help minimize costs and maintain reliability during the remainder of 2025.

2023-4 Align FFSS Pricing and Deployment Practices with Market Operations

The Firm Fuel Supply Service (FFSS) was implemented in 2022 in response to widespread generation shortfalls caused by fuel supply disruptions during Winter Storm Uri. FFSS provides a seasonal payment to selected generators in exchange for maintaining firm access to fuel during the winter months, from mid-November through mid-March. These generators, designated as Firm Fuel Supply Service Resources (FFSSRs), must be ready to operate if fuel supply disruptions threaten system reliability. FFSSRs can be deployed during declared winter weather watches; however, the specific criteria for deployment beyond this trigger remain unclear and are left to operator discretion.

ERCOT deployed FFSSRs over five consecutive days in January 2024 during a winter weather watch due to concerns about potential fuel supply disruptions. Across these five days, only one day exhibited reserve levels or pricing outcomes that indicated relatively tight conditions, and it remains unclear whether any observed impacts were directly caused by disruptions to the natural gas distribution network. The lack of clear and objective deployment criteria raises concerns about the efficient use of FFSS resources and the resulting costs imposed on the market.

The design of the FFSS program creates two significant market inefficiencies that increase costs for consumers. First, because FFSSRs receive upfront payments to secure fuel, they are incentivized to offer into the market at low prices regardless of the actual cost of burning expensive backup fuels such as fuel oil. This results in inefficient dispatch, where costly generation is selected even when less expensive alternatives are available. It also increases the cost of refueling these generators for future deployments and risks depleting firm fuel inventories prematurely. The costs associated with FFSSR operations, including the fixed readiness payments and any additional real-time costs, are ultimately spread across all LSEs through uplift charges.

Second, when FFSSRs are deployed, ERCOT removes their capacity from the calculation of online reserves used to determine the ORDC price adder. This artificial reduction in reported reserves raises scarcity prices even when physical capacity remains available. Given the lack of transparency around the operational rationale for deploying FFSSRs, this inflation of shortage pricing appears arbitrary and imposes unnecessary costs on the market. Although the upcoming transition to RTC should eliminate this practice by shifting reserve valuation into the ASDC framework, this issue will persist through November 2025 unless ERCOT takes interim action.

To address these concerns, we recommend that ERCOT:

1. Require FFSSRs to offer energy at prices that accurately reflect the true marginal cost of using firm fuel.
2. Include the capacity of deployed FFSSRs in the online reserves calculation until RTC is implemented.

2022-1 Implement a Multi-Interval Real-Time Market

ERCOT's real-time market currently only considers a single interval where dispatch instructions are optimized based only on forecasted system condition five minutes into the future. This limited forward-looking window limits the market's ability to efficiently dispatch and coordinate resources, particularly as the penetration of intermittent renewables and energy storage resources increases in ERCOT.

A MIRTM, which would optimize dispatch across at least six intervals (thirty minutes), offers several important benefits:

- It would allow more efficient scheduling of ESRs by considering the value of preserving or adjusting state of charge over multiple future intervals;
- It would better manage the utilization of NSRS, much of which is provided by offline resources, by enabling the real-time market to commit these thirty-minute reserves units more efficiently.
- It would help address the sharp evening net load ramp caused by increasing solar entry, often referred to as the "duck curve," by pre-positioning slow-ramping resources earlier and reducing unnecessary reliance on expensive ESRs and quick-start units.

Given these trends, we continue to recommend that ERCOT prioritize the implementation of a MIRTM following the completion of the RTC project.

2021-1 Eliminate the "Small Fish" Rule

Under the so-called "small fish" rule, generators with less than 5% of the installed capacity in ERCOT are deemed not to have "ERCOT-wide market power."[4] This rule was originally implemented before ERCOT adopted effective shortage pricing through the ORDC and was intended to allow smaller suppliers to submit offers significantly above marginal cost to help produce high prices during shortage conditions.

4 *See* 16 TAC § 25.504(c).

The introduction of the ORDC made this rule unnecessary. Shortage pricing now raises prices automatically based on reserve scarcity, meaning small suppliers no longer need to submit high offers to achieve efficient price signals. Despite this, economic withholding by small participants has led to instances of inefficient pricing. Withholding should not be permitted by any supplier that can be pivotal, and small entities can become pivotal during system-wide tight conditions or when the market is ramp constrained.

We recommend eliminating the small fish rule. PUC staff supported this recommendation in their response to last year's report and indicated they intend to open a project on this in 2025.

2021-2 Implement an Uncertainty Product

The rapid increase in intermittent renewable generation has introduced new operational reliability challenges stemming from forecast uncertainty. Over-forecasted renewable output can cause QSEs to under-commit thermal resources, leading to tight system conditions. ERCOT has responded by increasing its reliance on the RUC process and procuring larger quantities of 30-minute reserves, both of which distort market outcomes and impose excess costs through uplift. To address this issue, we recommend that ERCOT implement a longer-term reserve product provided by resources that can start within 2 hours when forecast uncertainties materialize.

In 2023, HB 1500 directed ERCOT to implement such a product, now referred to as DRRS. We recommend the following two design principles:

1. DRRS should be capable of being provided by both online and offline resources. Allowing both types of resources to participate ensures ERCOT can secure reserves at the lowest cost and avoids economically inefficient withholding from the energy market.

2. DRRS should be co-optimized with energy and other ancillary services in both the day-ahead and real-time markets, with shortages priced through a sloped demand curve. Co-optimization ensures that DRRS procurement reflects its marginal value, supports efficient price formation, and sends appropriate investment signals for resource adequacy.

If designed correctly, DRRS should reduce reliance on costly out-of-market actions and lower reserve procurement costs compared to holding excessive 30-minute reserves.

To date, ERCOT's DRRS proposals have not fully satisfied these criteria. A primary challenge stems from the statutory requirement that DRRS procurement offset RUC commitments. ERCOT has proposed deploying DRRS through the existing RUC process, but because RUC runs only once an hour and takes up to 30 minutes to complete, this approach complicates real-time co-optimization. Specifically, intra-hour SCED awards for DRRS cannot be incorporated quickly enough into RUC deployment decisions.

To address this, we propose that ERCOT develop a faster RUC process with a 4-5-hour study horizon. Our analysis suggests such a process could execute in as little as 5 minutes and could be run every 15 minutes or as needed at operator discretion. A faster RUC process would allow ERCOT to incorporate offline DRRS awards into commitment decisions in near real-time.

As the installed capacity of wind, solar, and storage resources grows, the reliability risks associated with renewable forecast errors will increase. Developing a fast, flexible unit commitment process for DRRS would support broader reliability needs while complying with HB 1500. We continue to recommend that ERCOT move forward with implementing DRRS using the design principles outlined above.

2020-3 Reconfigure Load Zones to Reflect Prevailing Congestion Patterns

Load zones are used to aggregate load nodes on the ERCOT grid that experience similar patterns of congestion. Unlike generators, which are settled at the nodal level and receive payments based on their specific locational marginal prices (LMPs), loads are billed based on the load zone price. This price reflects a load-weighted average of nodal prices within the zone. By aggregating prices in this way, the system shields loads from the full volatility of nodal pricing and makes market settlements more administratively manageable.

Beyond simplifying settlements, load zone prices also provide important short and long-term economic signals. When congestion patterns are accurately captured, load zone prices can inform consumption or hedging decisions for existing loads or can signal to prospective loads which locations are most cost-effective to site new demand. These price signals ultimately factor into other long-term infrastructure decisions such as transmission planning.

ERCOT's current four load zones – West, North, South, and Houston – were established in 2003. At that time, the geographic distribution of both load and generation, as well as the generation mix itself, looked very different than it does today. Industrial load has grown substantially in West Texas, driven by electrification of oil and gas operations and data center development, and residential load has sprawled far afield from historical population centers. Moreover, load is increasingly served by renewable resources located far away from load centers. As these patterns shift, it is increasingly important to re-evaluate whether the current load zone boundaries still reflect meaningful economic and operational groupings.

One indication that the current load zone groupings no longer adequately reflect congestion patterns is that congestion within several zones has grown more severe, especially in the South and West zones. The broad aggregation of nodal prices within each zone now conceals important differences in the cost of serving load. This disparity results in distorted pricing signals that can lead to inefficient consumption and siting decisions. For example, the Panhandle consistently experiences lower-than-average nodal prices due to a concentration of renewable generation. In contrast, the Permian Basin, part of the same load zone, faces significantly higher nodal prices driven by oil and gas-related demand growth. Yet both areas face the same zonal price for settlement, despite very different cost and congestion profiles.

We recommend that ERCOT re-evaluate the current load zone configuration to reflect these evolving congestion patterns. Any change to the configuration must take effect at least four years after it is approved, due to the structure of the CRR auction design. In Section IV, we present a potential methodology for reconfiguring load zones that would better align settlement prices with actual system conditions. To date, ERCOT has not indicated any intent to revisit or revise the current load zone structure.

2020-4 Implement a Point-to-Point Obligation Bid Fee

Point-to-point (PTP) obligations are financial instruments that allow market participants to take positions on incremental congestion between the day-ahead and real-time markets. PTP bids are submitted into the day-ahead market and are evaluated as part of the market optimization process. Over the past decade, there have been numerous delays in solving and posting the results of the day-ahead market. ERCOT has identified that a key contributor to these delays is the large and growing volume of PTP obligation bids. These delays are disruptive to the market and create unnecessary risk for participants who rely on timely information to guide real-time decisions.

Because ERCOT currently charges no fee to submit PTP bids, participants have no incentive to limit the number of bids they submit. Many of these bids are uncompetitive, making them unlikely to clear and of little value to the market in terms of liquidity or price formation. The inclusion of so many uncompetitive bids increases the complexity of the optimization problem and the time required for the market software to arrive at a solution.[5]

Applying a small bid fee to PTP obligation bids would discourage excessive and uncompetitive bid submissions without deterring legitimate hedging or trading strategies. By incentivizing participants to focus on bids with a meaningful likelihood of clearing, such a fee would streamline the market-solving process and reduce the frequency of day-ahead market delays. We recommend that ERCOT implement a small fee on day-ahead market PTP obligation bids. ERCOT has indicated a willingness to impose such a fee, although no NPRR has yet been submitted to move this change forward.

2019-2 Price Ancillary Services Based on the Shadow Price of Procuring Each Service

Clearing prices for ancillary services should be based on the shadow price of procuring each service. ERCOT's current practice of nesting multiple variations of operating reserves under a single ancillary service, setting limits on the procurement volume for each sub-product, and then clearing all of the sub-products at the same price violates this principle. For example, Responsive Reserve Service (RRS) effectively includes three separate products that are each cleared at the same price:

5 ERCOT's regression analysis can be found at http://www.ercot.com/calendar/2021/1/25/221086-WMWG.

1. Primary Frequency Response (PFR), an automatic response proportional to deviations in system frequency that can be provided by generators, energy storage resources, and controllable load resources;
2. Fast Frequency Response (FFR), a type of relay response that can be provided by batteries and NCLRs;
3. Under-Frequency Relay (UFR), a type of relay response provided by non-controllable load that trips loads offline in the event of significant drops in system frequency.

Out of the total procurement volume for RRS, a maximum of 450 MW can be provided by FFR, and only 60% can be provided by the sum of FFR and UFR. Because this second limit is usually a binding constraint, there is a surplus of offers for both types of relay response. However, because all of the products clear at the same price, this surplus does not have the effect of driving clearing prices down as one would expect in a well-functioning market.

To resolve this issue, ERCOT should set separate procurement volumes for PFR, FFR, and UFR and allow each product to clear at the shadow price associated with this constraint. If PFR is inherently preferable to FFR or UFR, then excess PFR offers should be substitutable for either UFR or FFR if it is offered at a lower price. This change in ancillary service methodology would result in prices that more accurately reflect the value of each product and offers that reflect the marginal cost of providing each product. Therefore, we recommend that the clearing price of all ancillary services be based on all the constraints used to procure the services.

2015-1 Modify the Allocation of Transmission Costs by Transitioning Away from the 4CP Method

The cost of expanding and maintaining the transmission system in ERCOT is allocated to consumers using the 4CP method. Under this methodology, each consumer's transmission tariff for the following year is based on their load ratio share during the 15-minute intervals with the highest monthly system-wide demand from June through September.[6] The 4CP method was approved in 1996 and was intended to allocate transmission costs to those who contributed most to system-wide peak demand. It was also designed to send a signal to consumers to reduce their load during peak periods, thus forestalling the need for new transmission investments. Whatever virtues this methodology may have had once upon a time, it no longer functions according to cost causation principles and fails to send efficient signals for new transmission investments.

Today, transmission upgrades are not primarily driven by peak demand. Instead, they are based on congestion patterns observed throughout the year and whether upgrades are needed to reliably serve evolving load patterns. The geography of growth has shifted: industrial load for oil and gas production and computing is growing at a rapid pace in West Texas, and residential load growth is sprawling into exurbs further afield from population centers. The cost of building

[6] 16 Tex. Admin. Code §25.192. Transmission Service Rates: http://www.puc.texas.gov/agency/rulesnlaws/subrules/electric/25.192/25.192.pdf

transmission now depends heavily on the distance between where generation is located and where load is growing. Connecting remote renewable generation to distant load centers requires significant investment in long-distance transmission infrastructure. The 4CP method, based solely on system-wide summer peaks, is inadequate to reflect the complex and location-specific drivers of today's transmission needs.

This misalignment is further exacerbated by the growth of LFLs, the majority of which are located in West Texas. These customers are well positioned to game the 4CP methodology by strategically curtailing their demand during intervals they anticipate could set a coincident peak. As a result, they impose significant transmission costs by locating in remote and less developed parts of the network but avoid paying their fair share of those costs. Because 4CP charges are based solely on contribution to peak demand, LFLs are able to shift a disproportionate share of transmission costs onto other consumers. This behavior undermines both the fairness and efficiency of transmission cost allocation.

We have recommended moving away from the 4CP method for more than a decade. During that time, billions of dollars have been spent on developing 4CP forecasting tools and installing behind-the-meter generation, creating vested interests in maintaining the current structure. However, these investments do not change the fact that 4CP no longer allocates costs in a way that reflects how the transmission system is built or used. Transmission cost allocation has recently drawn renewed legislative attention, with a number of pending legislative proposals in 89th Session calling for a review of the 4CP framework. Although no proposed legislation has become law as of this writing, we continue to recommend that policymakers transition away from 4CP toward a methodology that better reflects the true drivers of transmission investment and ensures that all consumers pay a fair share of system costs.

Recommendations Set for Retirement

These recommendations are expected to be removed from future editions of the State of the Market report, either because they have already been addressed by ERCOT rule changes, because they are no longer relevant, or because they have been incorporated into a new recommendation. For some of them, rule changes were approved several years ago but have still not been implemented. For these, we will continue to list them in this section in future editions of the report until they have been implemented.

2023-1 Adjust Shadow Price Caps to Reflect Constraint Reliability Risk

NPRR 1230 addressed this issue by creating a method to raise shadow price caps for specific constraints when necessary. A summary of the issue and recommendation follows. In the real-time market, shadow prices represent the marginal value of relieving a transmission constraint. A high shadow price signals that resolving the constraint is critical to maintaining system reliability, while a low shadow price indicates less urgency. Shadow prices influence how the dispatch model prioritizes generation across different locations on the grid. Shadow price caps for transmission constraints were fixed at predetermined levels. These caps were often set too low relative to the true reliability risk of constraint violations. As a result, the real-time market sometimes failed to dispatch sufficient local generation to manage flows within limits. When this occurred, ERCOT operators were forced to take manual actions outside of market processes to protect the system. These manual interventions often raised overall system costs and introduced inefficiencies into market pricing. Under the new framework, caps can be adjusted to reflect the actual reliability risk or operational cost of constraint violations.

2023-2 Revise Proxy Offers for IRRs Without Energy Bids

This issue is now incorporated into Recommendation 2024-4, which addresses offer requirements and proxy offer methodology for both ancillary services and energy. The current practice of inserting an administrative proxy offer curve for IRRs that fail to submit a complete offer can lead to inefficient dispatch outcomes, particularly in the presence of congestion. The original recommendation proposed that ERCOT adopt a proxy offer methodology for energy that supports the real-time market's ability to generate dispatch instructions that maintain system reliability without introducing unnecessary market distortions.

2022-3 Allow Transmission Reconfigurations for Economic Benefits

This recommendation was addressed by NPRR 1198, which was approved in July 2024 and is scheduled for implementation by the end of 2025 or early 2026 according to ERCOT Projects. Currently, ERCOT's approval processes only allow constraint management plans for reliability reasons.[7] However, there are times in which a transmission reconfiguration can relieve congestion without negatively affecting reliability.[8] Such plans should be developed and utilized. Both Midcontinent ISO (MISO) and Southwest Power Pool (SPP) are moving forward with this effort, though MISO is further along.[9]

2022-4 Change the Linear Ramp Period for ERS Summer Deployments to 3 Hours

This recommendation was addressed by NPRR 1006, which was approved in June 2020 but has still not been implemented. It is currently earmarked for implementation sometime in 2026 according to ERCOT Projects. In all summer ERS deployments to date, resources returned to pre-instruction levels within approximately three hours.[10] However, the current time value

7 A constraint management plan is a set of pre-defined manual transmission system actions, or automatic transmission system actions that do not constitute a Remedial Action Scheme, which are executed in response to system conditions to prevent or to resolve one or more thermal or non-thermal transmission security violations or to optimize the transmission system.

8 These are not post-contingency actions and so should have a negligible impact on the control room.

9 *See, e.g.*, https://cdn.misoenergy.org/20230228%20RSC%20Item%2006%20Reconfiguration%20for%20Congestion%20Cost%20Update628023.pdf.

10 https://www.ercot.com/files/docs/2022/09/13/DSWG%20-%20ERS%20event%20deployment%207-13-2022.pptx

parameter for returning to the pre-instruction level in the reliability deployment price adder calculation (an output of the SCED pricing run) was 4.5 hours. This difference artificially inflated the reliability deployment price adder. We recommended adjusting this parameter to 3 hours during summer hours.

2022-5 Change the Lookback Period for the ORDC Mean and Standard Deviation Calculations

Upon implementation of RTC at the end of 2025, the ORDC used in the current market design will be retired, and shortage pricing will be determined by ASDCs for each operating reserve product. Thus, we are retiring this recommendation but incorporating relevant aspects in our new recommendation on the design of the ASDCs in Recommendation 2024-1a. Please refer to the new recommendation for a discussion of the market issue the recommendation addresses.

2021-3 Reevaluate Net Metering at Certain Sites

We have reconsidered this recommendation and decided to withdraw it, as it primarily relates to our pre-existing recommendation to reconsider transmission cost allocation according to 4CP. Net metering of loads is mainly problematic insofar as it allows them to avoid paying for the transmission system while still giving them access to the transmission system when it suits them. Addressing transmission cost allocation, particularly for loads with net metering arrangements, remains a top priority.

I. Review of Real-Time Market Outcomes

The performance of the real-time market in ERCOT is essential for two primary reasons:

- **Coordinating Resource Dispatch and Managing Grid Reliability:** The real-time market schedules generators to produce energy based on system needs and transmission constraints. This ensures ERCOT maintains reliability across the grid while minimizing total production costs.
- **Setting Efficient Prices for Energy and Ancillary Services:** Real-time prices reflect the marginal cost of electricity at each location and moment in time. These prices incentivize generators to provide energy or reserves in the short term and guide long-term decisions about building new resources or retiring existing ones.

Only a small share of the power produced in ERCOT is settled in the real-time market. However, real-time energy prices set the expectations for prices in the day-ahead market and bilateral forward markets. Real-time prices are, therefore, the principal driver of prices in these markets where most transactions occur and inform long-term investment and retirement decisions.

In general, we have found that the real-time markets have performed well and produced prices that are competitive and efficient with the exception of limited periods adversely effected by the ERCOT Contingency Reserve Service (ECRS). In this chapter, we summarize and evaluate the outcomes of the ERCOT real-time market in 2024. We conclude with an examination of the extent to which the dysfunctions associated with ECRS continued to be prevalent in 2024.

A. Real-Time Market Prices

1. All-In Cost of Electricity

Figure 1 summarizes the "all-in" cost of supplying energy to serve load in the ERCOT wholesale market. In addition to the costs of energy, loads incur costs associated with ancillary services and a variety of non-market-based expenses referred to as "uplift." The all-in price metric includes the load-weighted average of the real-time market prices from all zones, as well as ancillary service costs and uplift costs divided by real-time load to show costs on a per MWh of load basis.[11] The energy prices are broken into three categories to show the effects of the two energy price adders on the overall cost of energy:

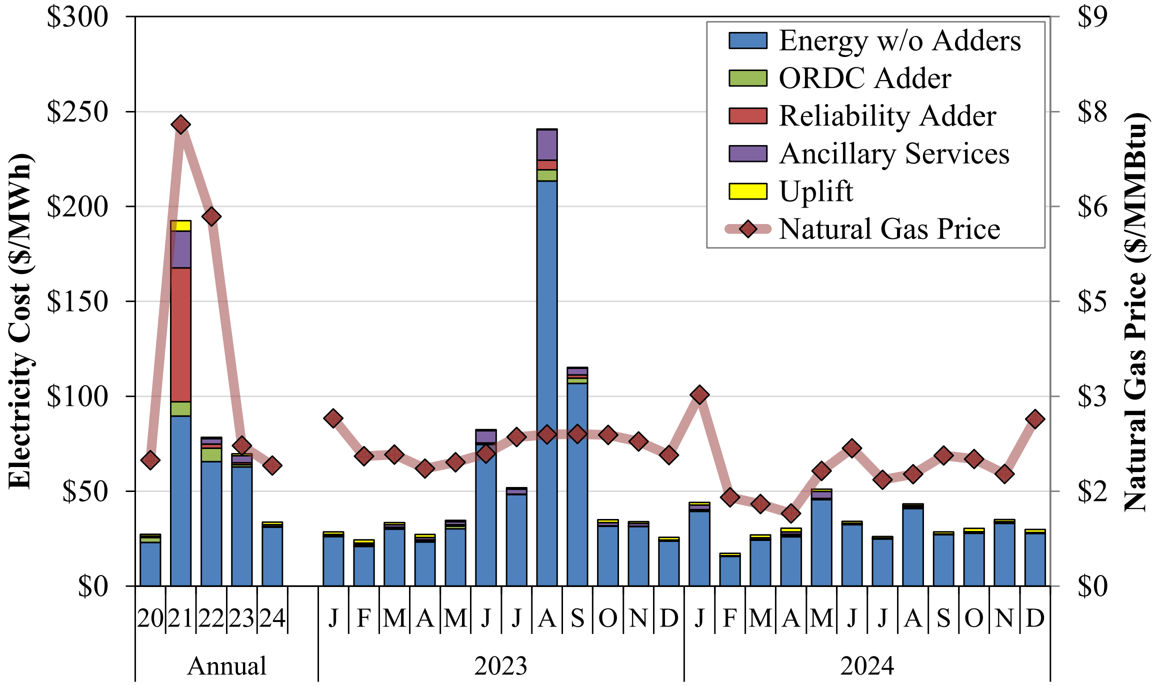
- The Operating Reserve Demand Curve (ORDC) Adder, implemented in 2014 to allow prices reflect the increasing reliability risks when reserves begin to run short; and

[11] For this analysis "uplift" includes: Reliability Adder Imbalance Settlement, ORDC Adder Imbalance Settlement, Revenue Neutrality Allocation, Emergency Energy Charges, Base Point Deviation Payments, ERS Settlement, Black Start Service Settlement, Block Load Transfer Settlement, Firm Fuel Service Settlement, High Dispatch Limit Override Settlement, RMR Settlement, RUC Settlement, Voltage Services Settlement, and the ERCOT System Administrative Fee.

- The Reliability Deployment Price Adder (RDPA), implemented in 2015 to ensure prices are not inefficiently reduced when ERCOT takes out-of-market reliability actions.[12]

These adders are the primary means for ERCOT to reflect shortage pricing through its markets. Figure 1 shows the monthly load-weighted average all-in prices for electricity in ERCOT the last two years and the annual average all-in prices for the last five years.

Figure 1: Average All-in Cost for Electricity in ERCOT, 2020-2024



The all-in cost of electricity fell roughly 52% from almost $70 per MWh in 2023 to $34 per MWh in 2024. This price drop was anticipated because of (1) decreased natural gas prices and (2) increased supply, particularly from solar and energy storage resources.

Correlation between gas price and energy price is expected in a well-functioning, competitive market because suppliers in a competitive market have the incentive to offer energy according to the marginal cost of generation. Fuel costs represent the largest component of the marginal production cost for most generators and natural gas is the most widely used fuel in ERCOT. However, the fact that energy prices remained at approximately the same level in 2023 as in 2022 despite a 62% drop in gas prices indicated that the real-time market was distorted in 2023 as the result of ECRS procurement and deployment practices.

The increase in available supply in 2024 played a significant role in reducing the observable impact of ECRS deployment practices on real-time market outcomes. In 2023, a large portion of

12 The reliability adder uses the dispatch software to simulate the system lambda without RUCs, deployed load capacity, or certain other reliability actions. The adder is the difference in system lambda output by SCED with and without any reliability actions.

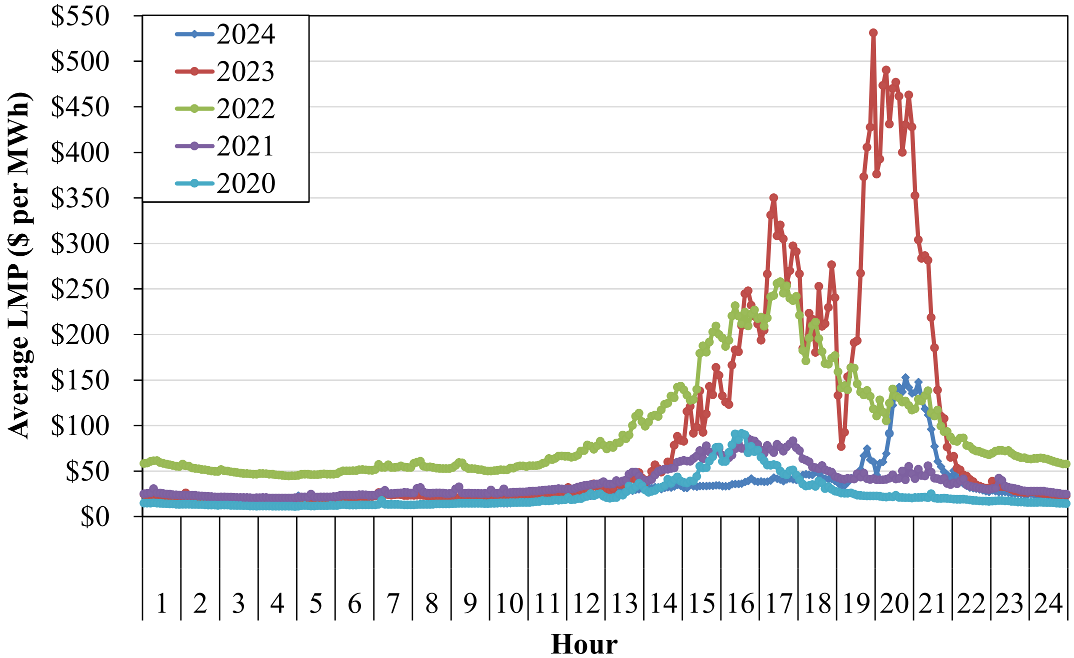
the costs associated with ECRS stemmed from artificial scarcity conditions perceived by the Security-Constrained Economic Dispatch (SCED) model. These conditions were less common in 2024 due to the substantial increase in online generation capacity. Figure 1 yields some additional insights:

- Ancillary services costs were $0.98 per MWh of load in 2024, a 74% decrease from 2023 and the lowest since 2020. We analyze this in greater detail in Chapter III.
- Uplift costs accounted for $1.24 per MWh of the all-in price in 2024, up 26% from 2023. Total uplift costs in 2024 were approximately $573 million. This was primarily due to the increase in the Real-Time Revenue Neutrality Allocation (RENA), the ERCOT System Administrative Fee, and costs for the Emergency Response Service (ERS) program.
- The increase in RENA of $161 million or $0.35 per MWh in 2024 can be attributed to differences between the load distribution factors (LDFs) used and transmission network modeling inconsistencies in day-ahead and real-time.

2. Prices by Time of Day

While Figure 1 shows the variation in the all-in-cost of energy on a monthly basis, Figure 2 illustrates how the real-time price of energy varies by time of day. Specifically, Figure 2 shows the load-weighted locational marginal prices (LMP) for energy for each 5-minute interval during the summer when demand and prices are typically the highest.

Figure 2: Prices by Time of Day, May-September 2020-2024



The price trends from 2020 to 2024 reflect the shifting net peak load hour from late afternoon (4 to 5 p.m.) to later evening hours (8 to 9 p.m.), which demonstrates the growing impact of solar generation. Solar reduces net load in the afternoon but ramps down quickly in the evening while demand is still high. As a result, the timing of peak prices has moved accordingly. The sharp difference in prices for 2024 compared to 2023 also highlights the significant effect of ECRS implementation artificially increasing the prevalence of shortage pricing.

3. Price Spike Impacts

To better observe the effect of the highest-priced hours on the average real-time energy price, Figure 3 shows the frequency of real-time energy price spikes in 2023 and 2024. For this analysis, price spikes are defined as 15-minute intervals when the load-weighted average energy price is greater than 18 million British thermal units (MMBtu) per MWh (i.e., an implied heat rate of 18) multiplied by the prevailing fuel index price (FIP) which produces an energy price spike threshold in $ per MWh. Prices at this level typically exceed the marginal costs of virtually all on-line generators. The figure also shows the portion of the average energy price in the month that is attributable to the price spikes.

Figure 3: Impact of Price Spikes on Real-Time Energy Price, 2023-2024

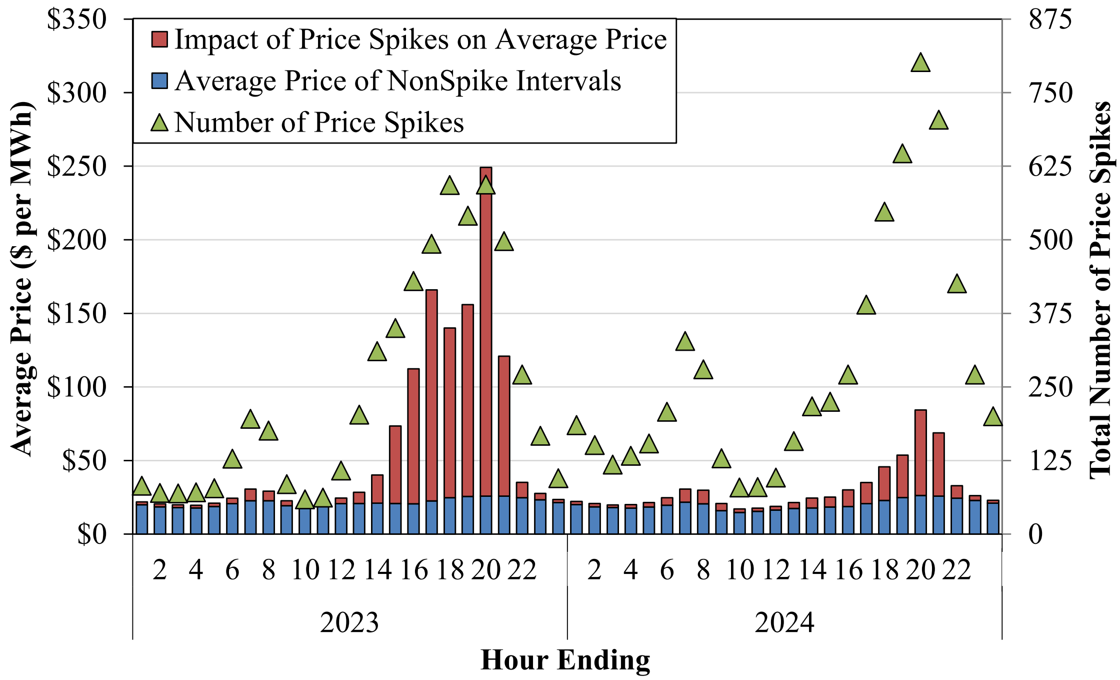


Figure 3 shows that price spikes were more frequent in 2024 than in 2023 but less impactful on the average price level. This trend is largely a result of the increase in solar generation, which lowers average prices in the daytime when demand is high but increases the likelihood of price spikes while the sun is going down and demand remains high. During these ramp-down periods, SCED commonly dispatches quick-start natural gas plants or energy storage resources with high priced energy offer curves, resulting in an increase in the frequency of price spikes despite having lower energy prices on average. Average prices were higher and the impact from price spikes was more significant in 2023 primarily due to more frequent extreme temperatures and the corresponding impacts of artificial scarcity caused by the implementation of ECRS.

B. Zonal Energy Prices

The cost to serve load varies by location because of congestion, and the result is differences in electricity prices between Load Zones. Table 1 lists the annual load-weighted average prices for each zone for 2020-2024.

Table 1: Average Annual Real-Time Energy Market Prices by Zone, 2020-2024

	2020	2021	2022	2023	2024
Energy Prices ($/MWh)					
ERCOT	**$25.73**	**$167.88**	**$74.92**	**$65.13**	**$31.91**
Houston	$24.54	$129.24	$81.07	$64.72	$29.58
North	$23.97	$206.39	$75.52	$68.55	$29.96
South	$26.63	$187.47	$72.96	$63.34	$34.34
West	$31.58	$105.27	$65.53	$61.62	$35.33

The most noteworthy change in 2024 is that average price in the West zone is the highest of the four zones for the first time since 2020, even though it had the lowest average price for 2021 through 2023. Moreover, the two highest zonal prices on average were in the West and the South for the first time since 2020.

The higher prices in the West and South are caused by greater exposure to congestion compared to the North and Houston zones, a topic we explore more thoroughly in Chapter IV. The West had the highest number of prices below $0, largely due to high wind and solar output that could not be delivered because of transmission congestion, both within the zone and on export paths to other regions. The West and South zones were tied for the greatest frequency of prices exceeding $200, driven by localized congestion within those zones.

C. Shortage and Reliability Pricing

The current ERCOT market design features two distinct price adders, the ORDC and the RDPA. This section summarizes the rationale and methodology for each of these price adders and reviews their impacts on real-time energy prices in 2024.

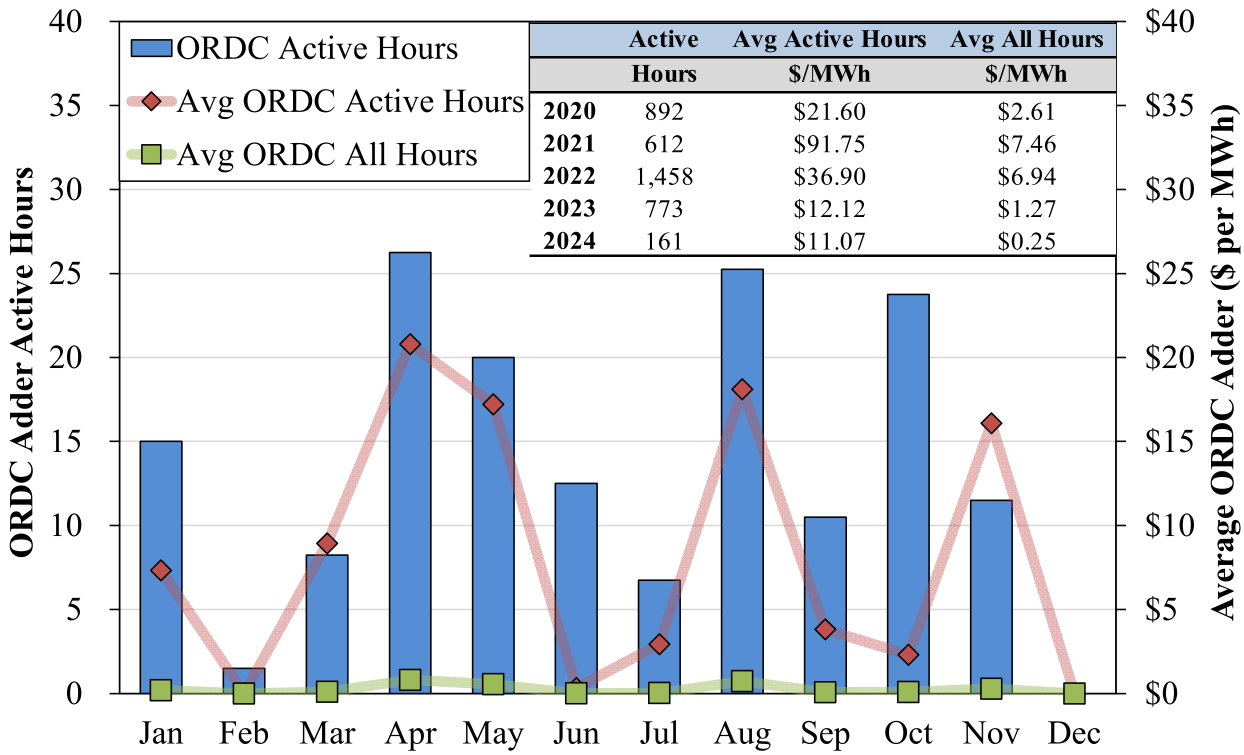
1. ORDC

The ORDC is intended to reflect the relationship between operating reserves and the probability of load shed, i.e., as operating reserves decrease, the likelihood of needing to shed load increases. The ORDC is formulated by comparing this relationship between the level of operating reserves and the probability of load shed to the Value of Lost Load (VOLL). Shortage pricing is considered efficient when this reliability cost is built into both energy and reserve prices during times of shortage. In ERCOT's current market design, this is done by adding an ORDC price adder to energy and reserve prices when reserves are low.

Over the years, ERCOT has made several changes to the ORDC to improve generator revenues and support reliability until the new Real-Time Co-optimization (RTC) system is in place. ERCOT summarized these changes and their impacts in their 2024 Biennial ERCOT Report on the Operating Reserve Demand Curve published in November 2024.[13] The most recent change to the ORDC was implemented in November 2023, when ERCOT implemented a multi-step ORDC price floor, the chosen "bridge solution" to increase the incentive to build new dispatchable generation until a reliability program could be implemented. In 2024, this price floor was binding for 247 SCED intervals and resulted in $11 million in additional revenue.

In 2024, the ORDC was active for only 161 hours,[14] just over one-fifth of the previous year's total. This sharp decline reflects the significant increase in generation capacity that began commercial operations in 2024, reducing instances of reserve shortages on the ERCOT grid. Additionally, a mild summer in 2024 kept demand lower, preventing grid conditions from tightening enough to trigger the ORDC as frequently. Overall, the ORDC contributed less than 1% of the annual average real-time energy price. Figure 4 summarizes the impact of the ORDC on real-time prices, showing the number of hours these mechanisms were active, their price impact during those hours, and their average impact normalized across all hours.

Figure 4: Impact of the ORDC on Real-Time Prices



	Active Hours	Avg Active Hours $/MWh	Avg All Hours $/MWh
2020	892	$21.60	$2.61
2021	612	$91.75	$7.46
2022	1,458	$36.90	$6.94
2023	773	$12.12	$1.27
2024	161	$11.07	$0.25

13 https://www.ercot.com/files/docs/2024/10/31/2024-biennial-ercot-report-on-the-ordc-20241031.pdf

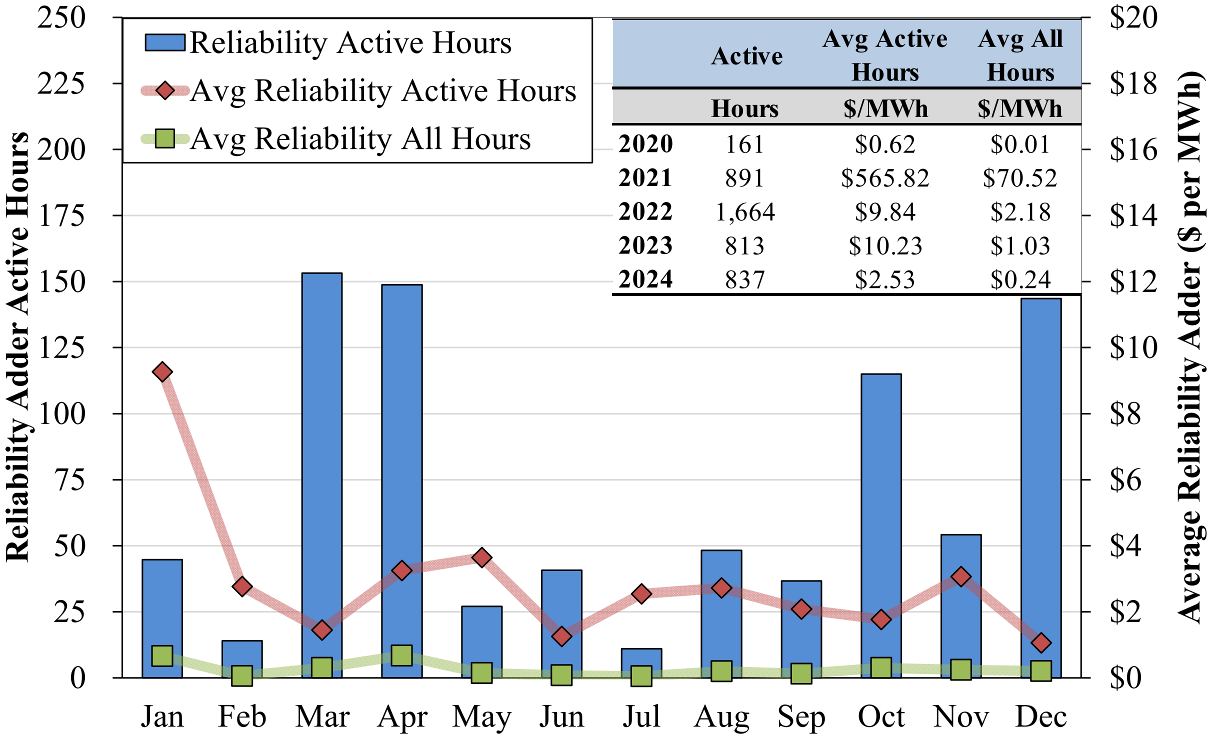
14 "Active" is defined as a settlement interval where the ORDC price was at least $0.01.

2. RDPA

The Reliability Deployment Price Adder is designed to offset the price-suppressing effects of out-of-market reliability actions taken by ERCOT. These actions, such as Reliability Unit Commitments (RUCs) and the deployment of ERS demand response, can artificially lower prices by increasing supply or reducing demand outside the normal market process. The RDPA corrects for this suppression by increasing real-time prices when these actions occur, ensuring that prices continue to reflect true market conditions and reliability risks.

The RDPA was triggered in roughly the same number of hours in 2024 as in 2023, but its average impact on real-time prices was less than a quarter of the previous year's. The explanation for this discrepancy is that while out-of-market reliability actions occurred just as frequently, they occurred under less extreme conditions in 2024 than in 2023, resulting in a lower impact on prices. For example, there were no ERS deployments in 2024, but ERS deployments on September 6, 2023, contributed to reliability pricing over $1,000 per MWh for several intervals, where the RDPA never exceeded $175 per MWh in 2024. Figure 5 summarizes these impacts, showing both the frequency of RDPA activations and their effect on prices.

Figure 5: Impact of the RDPA on Real-Time Prices



	Active	Avg Active Hours	Avg All Hours
	Hours	$/MWh	$/MWh
2020	161	$0.62	$0.01
2021	891	$565.82	$70.52
2022	1,664	$9.84	$2.18
2023	813	$10.23	$1.03
2024	837	$2.53	$0.24

D. Aggregated Offer Curves

The next analysis compares the quantity and price of generation offered in 2024 and 2023. Figure 6 provides the average aggregated generator offer stacks for the year in all hours, the peak load hour, and peak net load hour of the year.

This figure shows that:

- 36% of real-time generation capacity was not dispatchable because it is below generators' Low Sustained Limit (LSL).
- 32% of real-time generation capacity was offered at less than or equal to $0, primarily from wind and solar resources. These resources have the incentive to produce even when prices are negative because many of them receive federal production tax credits.
- 16% of the capacity was priced at levels between zero and a value equal to 10 times the daily natural gas price (known as the FIP). This price range represents the incremental fuel price for the vast majority of the ERCOT generation fleet.
- Roughly 17% of the capacity was offered above this level in 2023. Note that $75 corresponds to the energy offer floor for capacity providing online Non-Spin Reserve Service (NSRS), which averaged approximately 1,000 MW across all of 2024.

Figure 6: Aggregated Generation Offer Stack – Annual, Peak and Net Peak Load

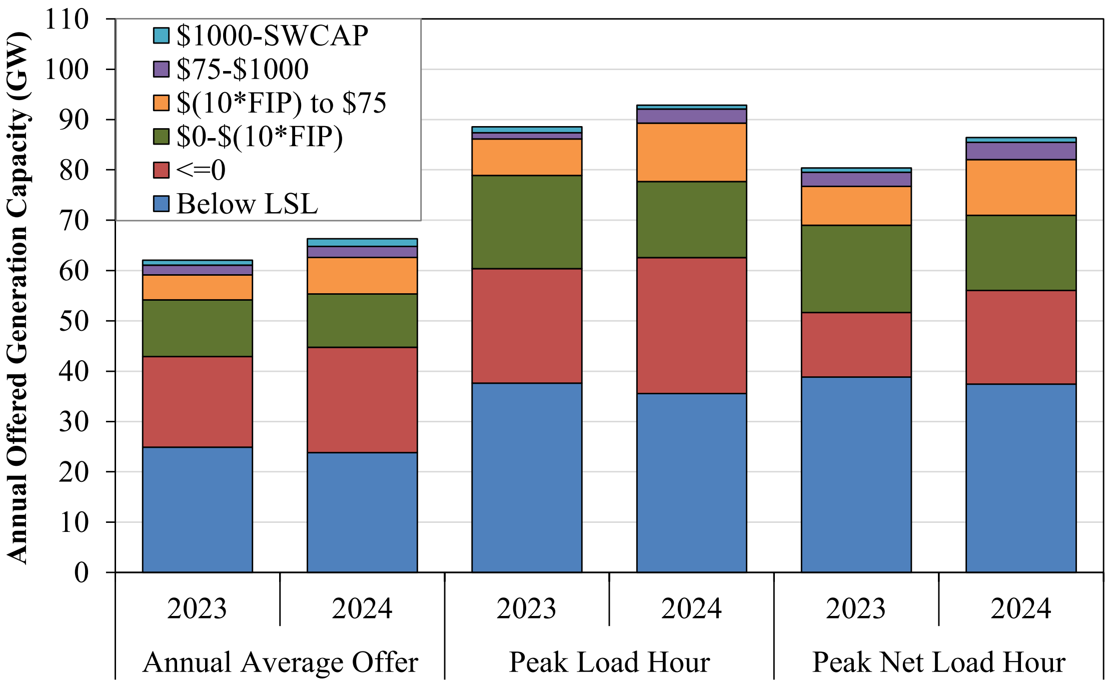


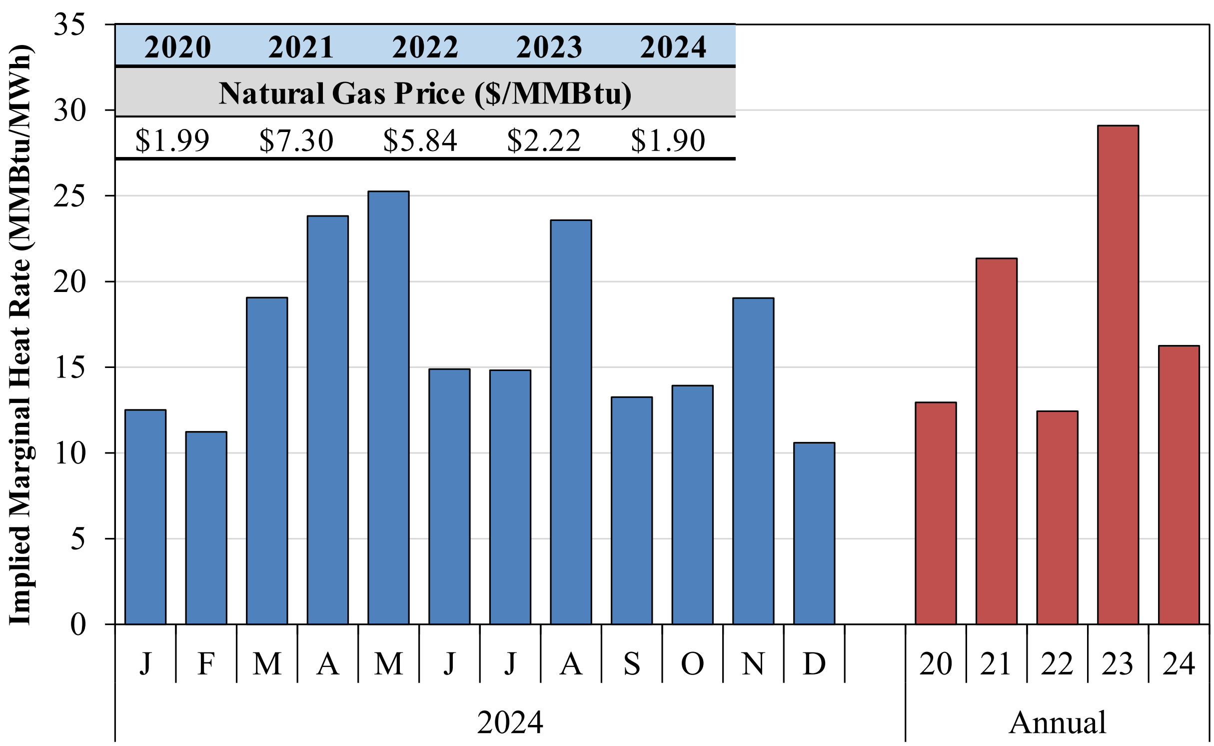
Figure 6 shows that the average amount of capacity above LSL that was offered into the real-time market in 2024 increased by approximately 5,300 MW on average compared to 2023. Nearly 3,000 MW of this increase can be attributed to offers at or below $0, mostly due to the large increase in installed solar capacity in 2024. Conversely, the volume of capacity below LSL decreased by 1,100 MW, indicating a decrease in committed capacity from thermal resources. This trend is also explained by the large increase in solar capacity that had the effect of decreasing pricing throughout the middle of the day, thus decreasing the economic opportunities for thermal commitments.

E. Real-Time Prices Adjusted for Fuel Price Changes

Historically, the real-time price of electricity was directly correlated with the price of fuel for the marginal generation technology, which for ERCOT is most commonly natural gas. Over time, other factors have weakened the connection between the prices of natural gas and electricity. Figure 6 highlighted that the significant deployment of zero marginal cost generation like wind and solar have resulted in hundreds of hours prices at or below $0 per MWh. Conversely, shortage pricing induced by the ORDC results in hundreds of hours per year of prices higher than the marginal cost of natural gas generators.

To summarize the separation between electricity and natural gas prices, we produce an "implied marginal heat rate" that is calculated by dividing the real-time price of electricity by the natural gas price. Figure 7 shows the implied marginal heat rates for each month of 2024 and on an annual basis for 2020-2024.

Figure 7: Monthly and Annual Implied Heat Rates, 2020-2024



The average implied heat rate for 2024 was considerably lower than in 2021 and 2023, years in which prices were significantly elevated due to Winter Storm Uri and ECRS-related issues. That said, the implied heat rate for 2024 was still roughly 30% higher than in 2020 and 2022 and noticeably higher than would be expected based on the marginal heat rate of the natural gas power plants in the ERCOT market. This disparity is driven by the impact from price spikes, shown in Figure 3, which were more impactful in 2024 relative to the price of fuel compared to 2020 and 2022.

F. Exposure of Load to Real-Time Prices

Although real-time prices play a critical role in sending investment signals in ERCOT's energy-only market, most load is not directly exposed to these prices. In 2024, only a relatively small share of total load was settled at real-time energy prices. Instead, many load-serving entities (LSEs) manage price exposure through a variety of mechanisms, including owning generation, entering into bilateral power purchase agreements (PPAs), participating in futures markets, or procuring energy through the day-ahead market. Figure 8 shows the percentage of load that was exposed to real-time prices based on ERCOT settlement data.

Figure 8: Monthly Average Load Exposure, 2020-2024

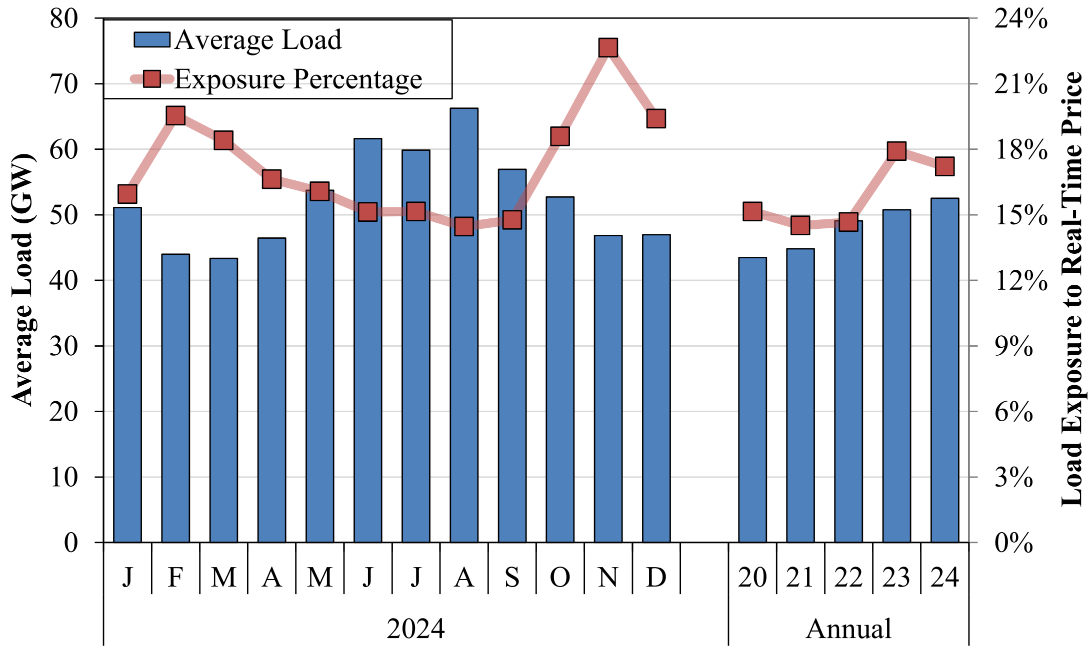


Figure 8 shows that only 17% of the load was exposed to real-time prices on average in 2024, and the remainder of real-time load is hedged with positions from the day-ahead market, including PTPs. This type of forward hedging provides some protection against volatility and uncertainty associated with the real-time market, but it ultimately provides little protection against sustained elevations in prices because day-ahead prices in a well-performing market will tend to reflect the expected real-time prices. For more on the convergence between day-ahead and real-time market pricing, see Chapter III.

G. Impact of ECRS on Real-Time Market Prices

The implementation of ECRS in June 2023 had a profound impact on the ERCOT wholesale market and is referenced throughout this report. The pricing impact of ECRS in 2024 was not as severe as it was in 2023, largely due to the fundamentals of supply and demand. There was significantly more solar and energy storage capacity online throughout 2024 than in 2023, and there were fewer days of extreme heat in the summer. The net effect was a marked reduction in hours with reserve shortages in 2024 relative to 2023. That said, problems remain with how much ECRS is procured and how it is deployed. This section includes a summary and analysis of ECRS procurement and deployment practices in ERCOT and their resulting impacts on the functioning of the real-time market.

1. Changes in Operating Reserve Procurements

The primary reason the ECRS rollout proved costly in 2023 was that it increased the amount of reserve capacity that SCED could not access for energy dispatch. In the absence of real-time co-optimization between energy and ancillary services, ERCOT must preserve ancillary service capacity by preventing SCED from assigning energy base points to that capacity. Ancillary service deployments, in this framework, effectively release the reserved capacity to SCED so it can be used for energy dispatch. Prior to ECRS, this limitation applied only to regulation and responsive reserves. NSRS were treated differently as they had the ability to be available to SCED with an energy offer floor of $75 per MWh. ECRS, once introduced, was withheld from SCED in the same way as regulation and responsive reserves, reducing overall dispatch flexibility and contributing to higher market costs.

Figure 9 shows ERCOT's average 10-minute and 30-minute reserve procurements from 2020 to 2023 compared to those of other Regional Transmission Organizations (RTOs). Figure 9 aggregates Responsive Reserve Service (RRS) and ECRS because they are both 10-minute reserve products, and NSRS are shown as 30-minute reserves.

Figure 9: Increase in Ancillary Services Procurement

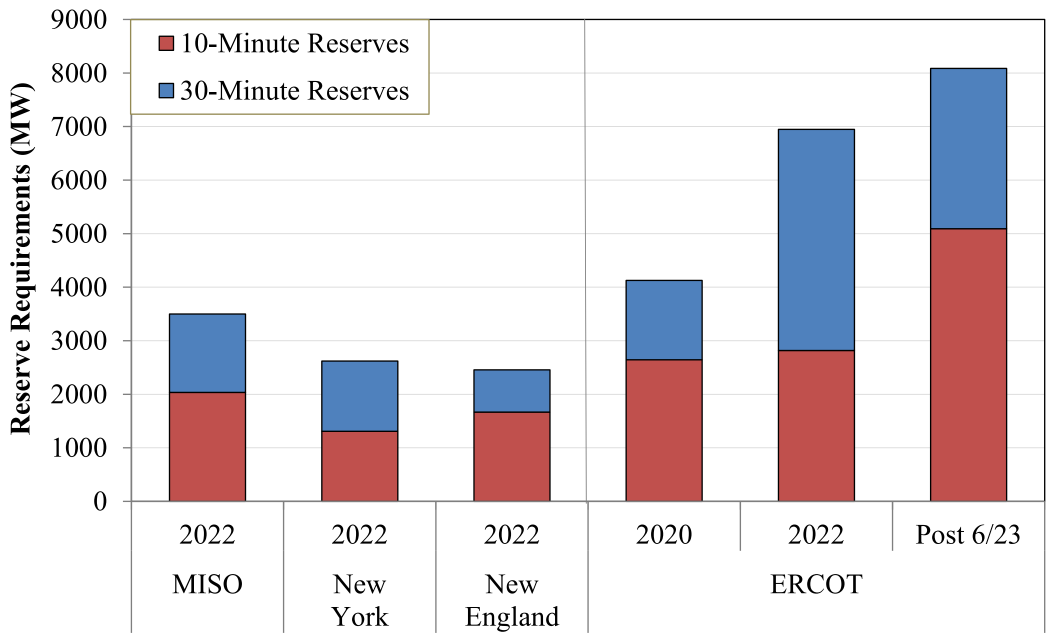


Figure 9 shows that in 2020, ERCOT procured higher levels of operating reserves than the levels procured by other RTOs, which is understandable because ERCOT is effectively an electrical island that must have the ability to respond to system contingencies quickly and effectively. The large increase in 30-minute reserve procurements in 2022 as ERCOT began conservative operations generated much higher reserve procurement costs. However, they did not significantly affect the real-time market outcomes since these reserves are not withheld from SCED. With the implementation of ECRS in 2023, however, the level of 10-minute reserves was substantially increased, significantly increasing the volume of operating reserves that are inaccessible to SCED. Next, we present our analysis of the ECRS procurement levels.

2. Assessment of ECRS Procurement Quantities

Ideally, operating reserve procurements should be aligned with the reliability objectives the reserves are intended to address, which we evaluate in this section. We used a stochastic risk model to characterize the relationship between ECRS volumes and the probability of firm load shed based on historical operating conditions. There are two primary reliability risks that ECRS address: (1) forced generation outages, where reserves are deployed to offset the loss of supply, and (2) net load forecast errors, which can lead to under-commitment of thermal resources.

We used ERCOT data on the historical rates and magnitudes of forced outages and forecast errors to determine the probabilities of these risks, then applied these risk probabilities to historic operating conditions to estimate the expected value of load at risk of being shed – this indicates the amount of ECRS needed to avoid outages. Using these results, we calculated how lowering the ECRS procurements would affect frequency and potential for load shedding as measured by an annualized Loss of Load Probability (LOLP). This relationship is shown in Figure 10.

Figure 10: Annual LOLP vs. Decrease in ECRS Procurement Volume

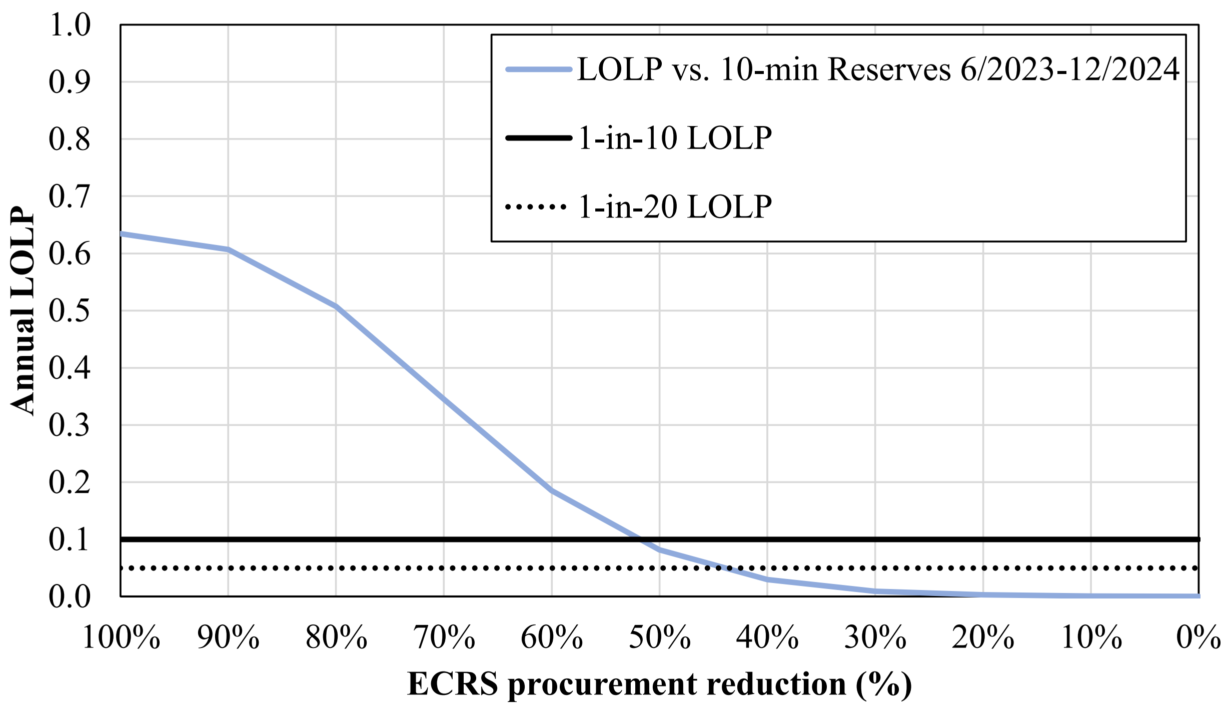


Figure 10 indicates that ECRS procurement volumes could be cut in half while maintaining an annual LOLP between 5-10%. Hence, ERCOT could mitigate the adverse market impacts of excess ECRS procurements while maintaining a high level of reliability. Next, we discuss the role of ECRS deployment practices in exacerbating these market performance issues.

3. Effects of ECRS Deployments on Real-Time Energy Prices

Upon the implementation of RTC, SCED will be able to make economic trade-offs between procuring energy and ancillary services in real-time with reserve shortages prices according to a corresponding ancillary service demand curve (ASDC). Until then, the functioning of the real-time market depends on judicious deployments of ancillary services. Our analysis in the 2023 State of the Market report indicated that ECRS deployment practice resulted in approximately $12 billion in excess costs in the real-time market.

In response to this analysis, ERCOT sponsored NPRR 1224[15] in the spring of 2024. This NPRR would codify conditions under which blocks of ECRS capacity could be released to SCED. The final proposal was that 500 MW of ECRS capacity would be released after two consecutive SCED intervals of under-gen[16] of at least 40 MW, effectively forcing the real-time market into artificial shortage conditions before releasing ECRS. Moreover, the capacity released to SCED would be subject to a $750 per MWh energy offer floor, for which there is no reasonable basis. Hence, we found this proposal to be flawed, and it was ultimately rejected by the Commission in July 2024. However, ERCOT still implemented the deployment process proposed by NPRR 1224, but without the $750 per MWh energy offer floor.

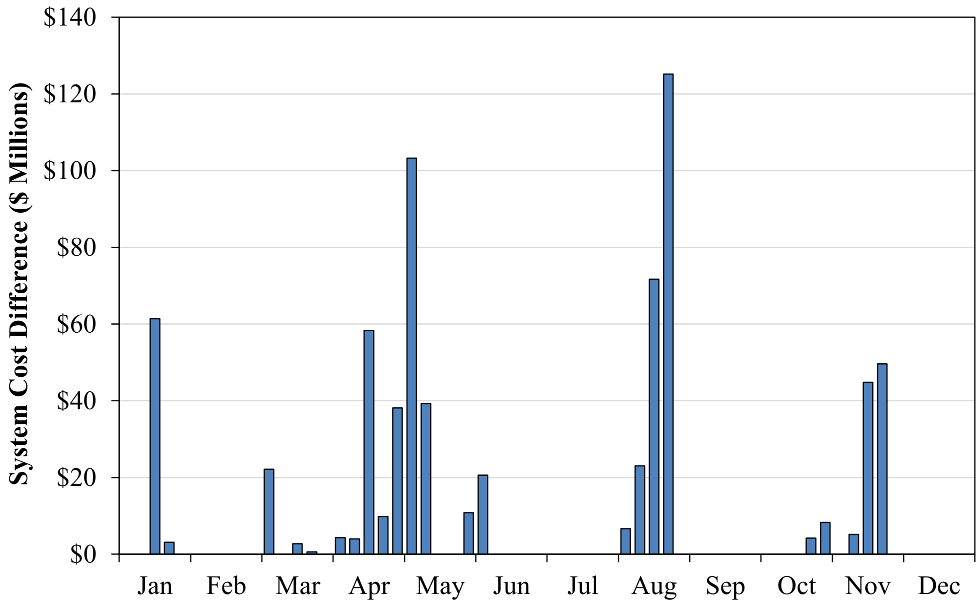
To estimate the excess costs caused by ECRS deployment practices in 2024, we identified real-time market conditions where deploying ECRS would have significantly reduced costs. We focused on events with at least two consecutive SCED intervals showing 10 MW or more of under-generation. These are situations where deploying ECRS would have clearly improved both economic efficiency and reliability. Importantly, deploying the ECRS does not generally reduce the 10-minute reserves available to ERCOT but simply makes them available to SCED and transfers them to online resources. In fact, it often increases available reserves by preventing storage resources from having to discharge to satisfy demand.

For each event, we measured its duration based on how long the under-generation continued. We then simulated these events assuming full release of ECRS capacity to SCED and compared the outcomes to actual SCED results. The excess cost was calculated as the difference between the original and simulated results, measured as system lambda multiplied by load. Figure 11 presents the estimated excess costs for each of these identified events.

15 https://www.ercot.com/mktrules/issues/NPRR1224

16 SCED generally procures enough energy through base points to satisfy demand, but it can go short on energy according to the Power Balance Penalty Curve. The volume of shortage is referred to as "Under-generation."

Figure 11: Excess Cost of ECRS Deployment Practice, 2024



In 2024, ECRS deployment practices contributed to both significant market costs and reliability challenges. Figure 11 identifies 24 events where more flexible deployment of ECRS would have saved the market over $700 million in total, two events exceeding $100 million in excess costs. Only three of these events included ECRS deployments according to the criteria defined by NPRR 1224, and the sum of excess cost from those three events was almost $220 million. These costs are much lower than the $12 billion in 2023 because the system was less tight.

In addition to the financial impact, current practices negatively affected reliability by excessively dispatching energy storage resources (ESRs) for energy, while keeping gas turbines in reserve. Simulations that released ECRS in these events preserved ESR state of charge (SOC) by reducing net injections by ESRs by 13% throughout the events.

4. ECRS Conclusions and Recommendation

ERCOT's current approach to procuring and deploying ECRS continues to create significant challenges for both market efficiency and system reliability. However, the implementation of RTC in late 2025 will address many of these issues. With RTC, SCED will have the ability to go short on ECRS based on the shortage prices defined in the ASDC, eliminating the need for manual deployments and enabling more efficient, market-driven decisions.[17] Until then, we continue to recommend that the ERCOT improve its procedures to release earlier before allowing consecutive intervals of artificial shortage pricing in the real-time market and to maintain the released ECRS until operating conditions have been resolved. These improvements would improve reliability and lower costs substantially.

17 The ASDCs defined by NPRR 1268 sponsored by the IMM, should result in further improvements in market performance vis a vis ECRS. See https://www.ercot.com/mktrules/issues/NPRR1268 for more details.

II. Demand and Supply in ERCOT

Many of the trends in market outcomes described in Chapter I are attributable to changes in the supply portfolio or load patterns in 2024. This chapter summarizes the trends in supply and demand including wind and solar generation, the dramatic increase in market participation from energy storage resources, and the behavior of demand response resources and other price responsive loads.

A. ERCOT Load in 2024

Figure 12 shows peak load and average load by zone from 2020 through 2024.[18] The average load characterizes the aggregate change in energy demand over the entire year, while the peak demand reflects the instantaneous demand for available generation capacity to avoid load shed.[19]

Figure 12: Annual Load Statistics by Zone, 2020-2024

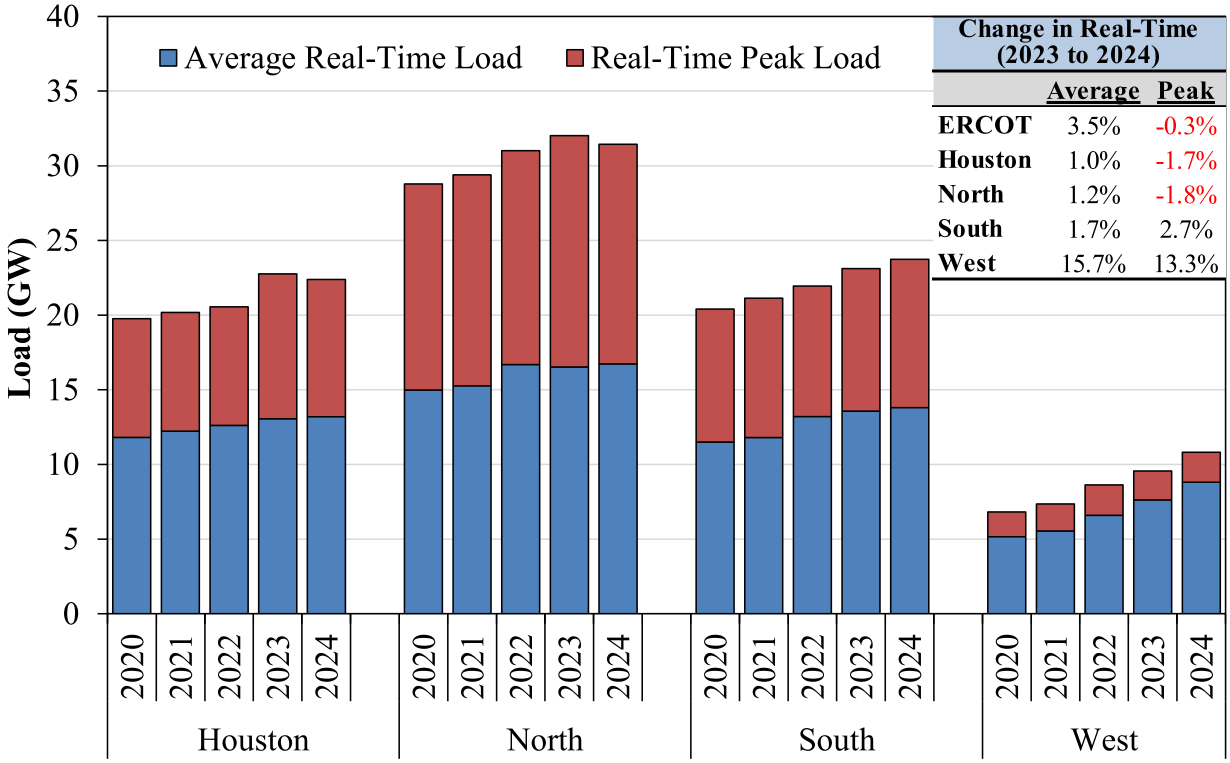


Figure 12 shows that the average ERCOT load in 2024 increased by 3.5%, but the peak load slightly declined. This disparity is explained by the continued population and economic growth in Texas coincident with less extreme temperatures than in 2023. The West continued to exhibit the largest relative increase in both average and peak load, 15.7% and 13.3%, respectively,

18 Non-Opt-In Entity (NOIE) load zones have been included with the proximate geographic zone.

19 In recent years, peak net load (load minus intermittent renewable output) is a more direct cause of shortages.

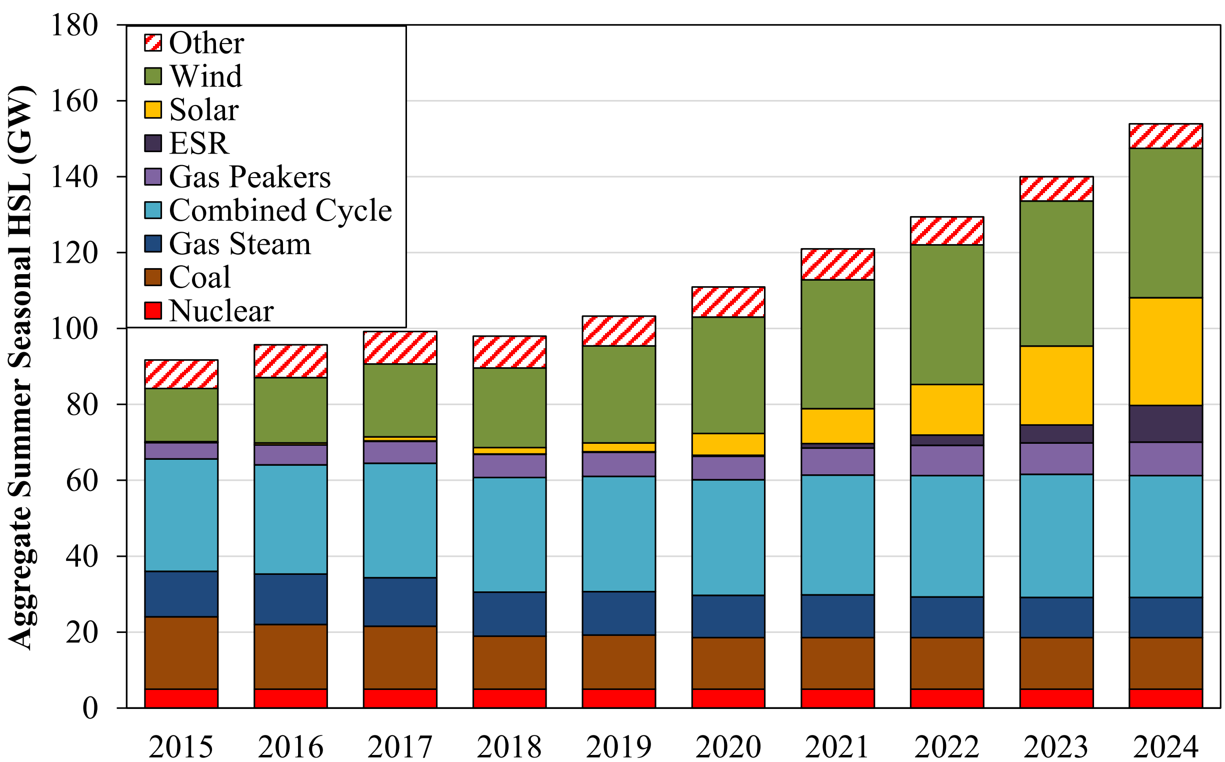
which is driven by the growth in oil and gas production in the Permian Basin and data center load related to cryptocurrency mining, which is discussed in further detail below in the Large Flexible Load Section.

B. Generation Capacity in ERCOT

ERCOT's installed generation capacity at the end of 2024 reflects continued growth in solar and storage, with most new additions concentrated in the South and North load zones. This section provides a breakdown of installed capacity by resource type, highlights how much came online in 2024, and identifies where those additions occurred across the system. Figure 13 shows total installed capacity with Part 2 Approvals to operate at their summer High Sustained Limit (HSL) from 2015 through 2024. It includes full HSL capacity for intermittent renewable resources (IRRs) and only reflects resources with HSLs telemetered from private use networks (PUN) that were available for ERCOT dispatch within the year.

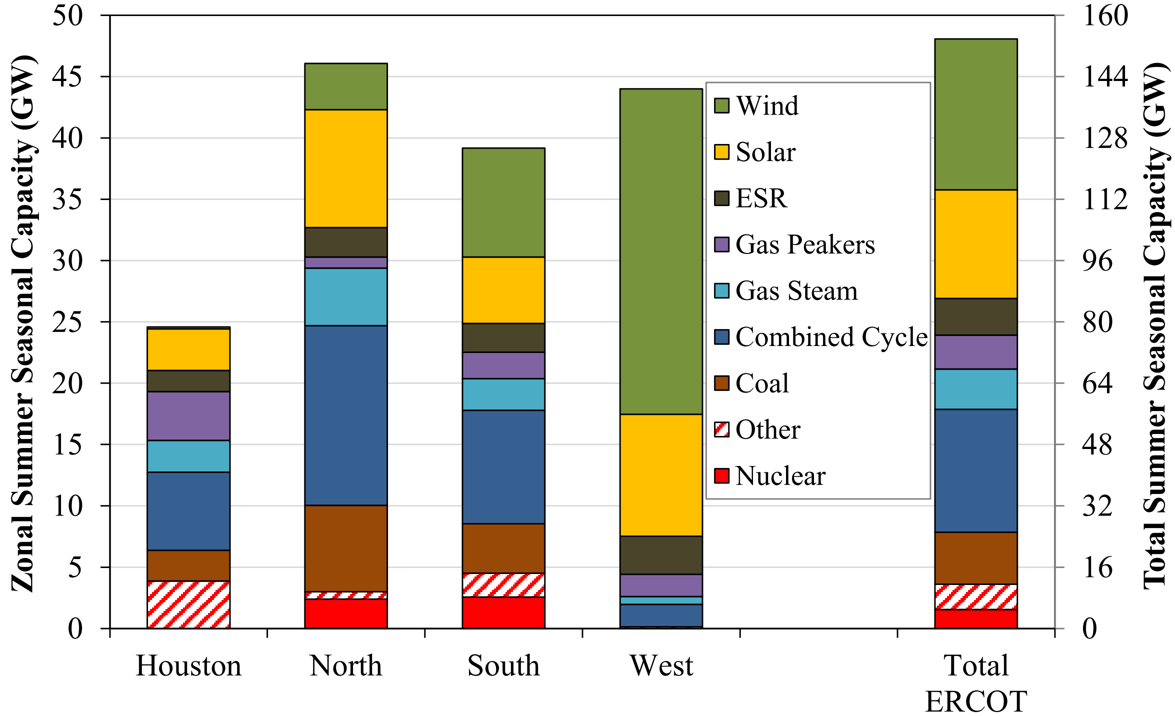
In 2024, approximately 14 GW of new capacity entered commercial operation, including 7.5 GW of solar, 5.0 GW of energy storage resources (ESRs), 1.1 GW of wind, and 500 MW of combustion turbines. Since 2020, most new capacity has come from solar and ESRs, with annual average additions of 5.2 GW and 2.4 GW, respectively.

Figure 13: Installed Generation Capacity in ERCOT, 2015-2024



Most of the new generation and storage capacity added in 2024 was built in the South and North zones, which accounted for 37% and 33% of new builds, respectively. Nearly all of this capacity was solar and storage. An additional 21% was built in the West zone, while the remaining 9% was added in the Houston zone. Figure 14 shows the total installed capacity by resource type across each zone.

Figure 14: Installed Capacity by Resource Type for Each Zone, 2024



The geographic distribution of capacity in the North and South zones closely mirrors the pattern of demand shown in Figure 12. In contrast, the West zone continues to generate more power than it consumes and remains a major exporter. The Houston zone has become increasingly dependent on imports from other parts of the state as Houston's load growth outpaced local capacity additions.

The composition of generation output has changed consistently with the changing installed capacity shown in Figure 13. Over the past year, the share of generation from wind has been nearly flat while the solar share increased from 7.2% in 2023 to 10.4% in 2024. The share of output from coal and natural gas resources both fell roughly one percentage point each from last year to 12.6% and 44.2% of all output in 2024, respectively.

C. Wind and Solar Output in ERCOT

The output of wind and solar resources has been growing over time and can vary substantially by season. To show these trends, Figure 15 and Figure 16 show monthly wind and solar generation totals for 2024 and annual generation from 2020 to 2024. They also show the total amount of wind and solar curtailment due to congestion. Wind generation grew by 3.9% from 2023 to 2024, with more than 4% of output curtailed due to congestion. Solar generation grew more sharply, increasing by 62% with nearly 6% of that output curtailed.

Figure 15: Wind Production and Curtailment, 2018-2024

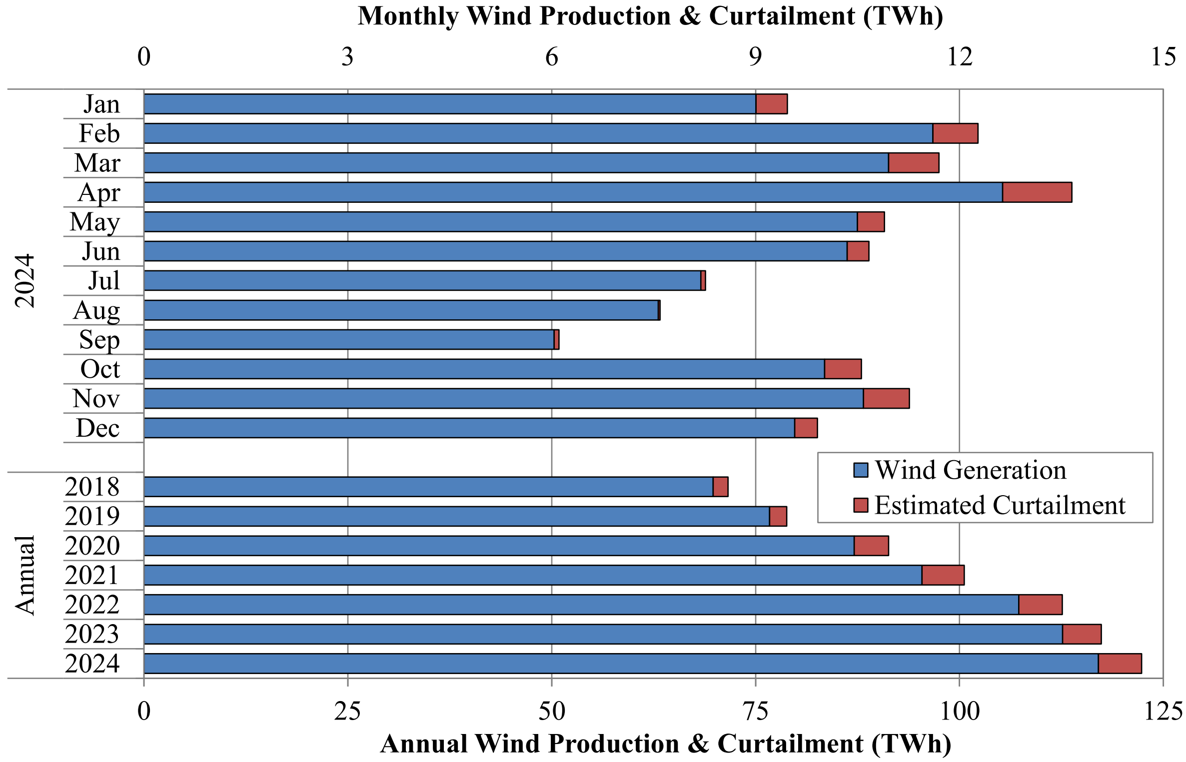
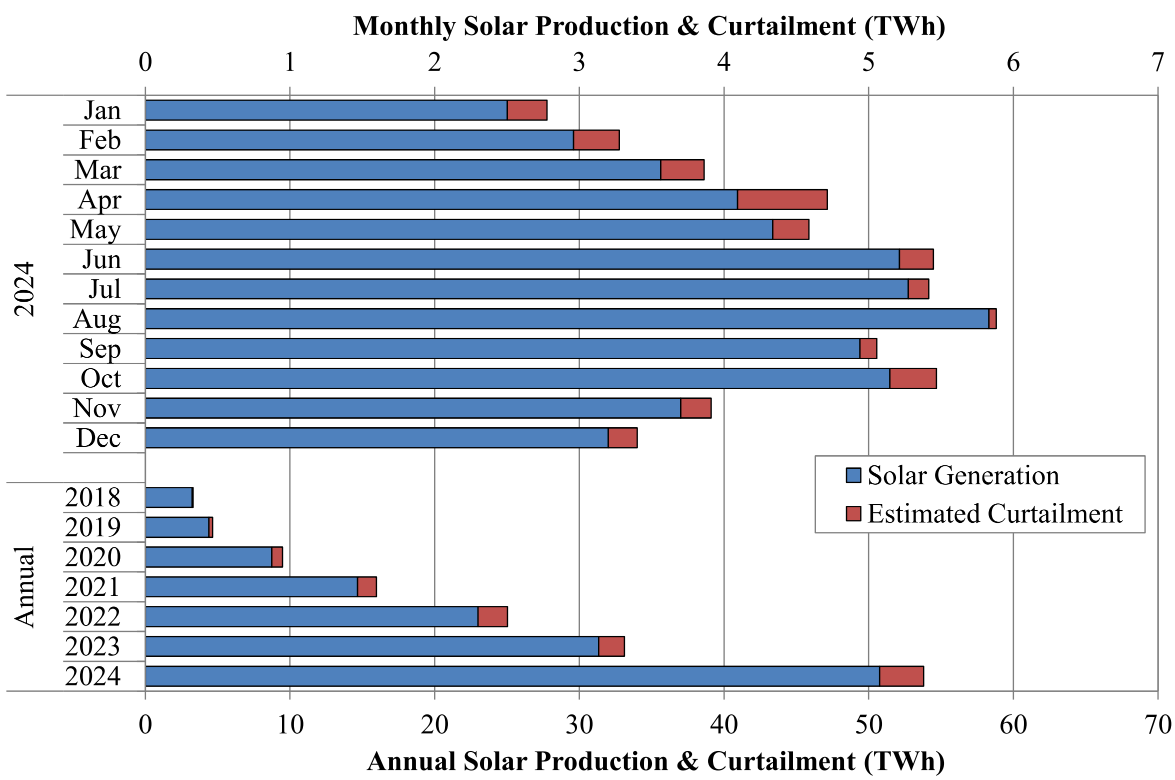


Figure 16: Solar Production and Curtailment, 2018-2024



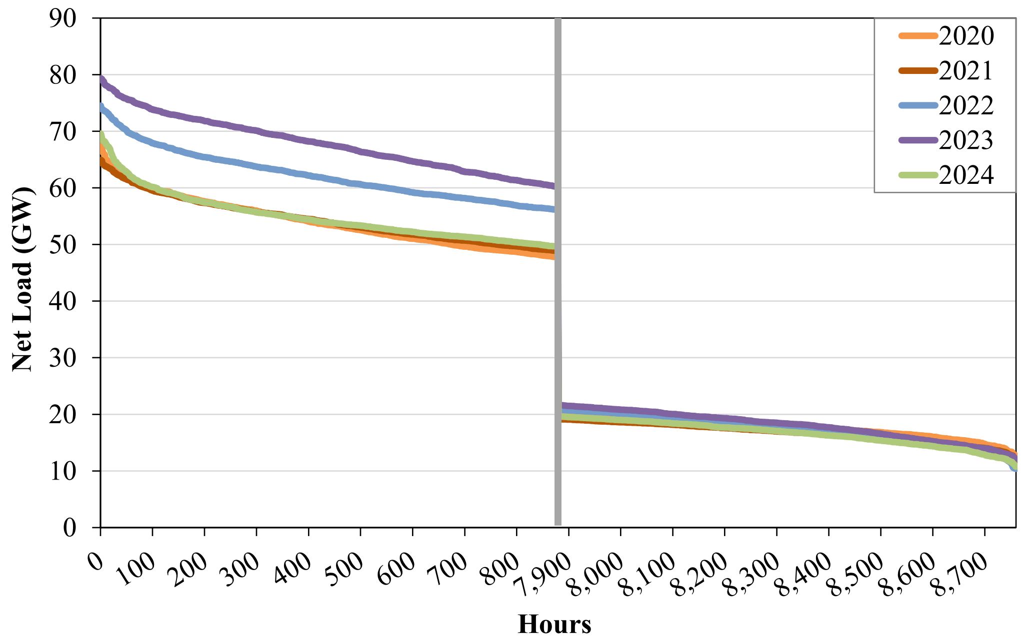
While wind has a higher overall capacity factor than solar, it is less aligned with daily peak load.[20] As shown in Table 2, wind resources had a 29% capacity factor during daily peak load intervals in 2024, compared to over 40% for solar. However, as solar output has increased, it has shifted the timing of peak net load, which is now occurring later in the evening when solar production declines. As a result, the solar capacity factor during the daily net peak net load interval has dropped each year since 2020, reaching just 3.9% in 2024. In contrast, wind maintained a capacity factor of 28.4% during the same interval, consistent with previous years.

Table 2: Aggregate Capacity Factor of Wind and Solar Generation, 2020-2024

	Wind			Solar		
Year	Overall	Daily Peak Load	Daily Net Peak Load	Overall	Daily Peak Load	Daily Net Peak Load
2020	35.3%	31.8%	26.2%	25.6%	44.8%	42.3%
2021	33.4%	30.6%	25.8%	23.6%	41.7%	33.1%
2022	34.4%	31.8%	28.2%	24.2%	40.9%	23.1%
2023	32.4%	28.9%	27.1%	22.7%	39.4%	11.7%
2024	32.0%	29.3%	28.4%	22.3%	40.1%	3.9%

As more energy is produced from renewables, especially during sunny or windy periods, net load becomes more variable and concentrated into fewer hours of the day. This shift affects both real-time system operations and long-term resource adequacy planning in ERCOT. Figure 17 highlights this trend by showing net load during the highest and lowest hours of 2024.

Figure 17: Top and Bottom Deciles (Hours) of Net Load, 2020-2024



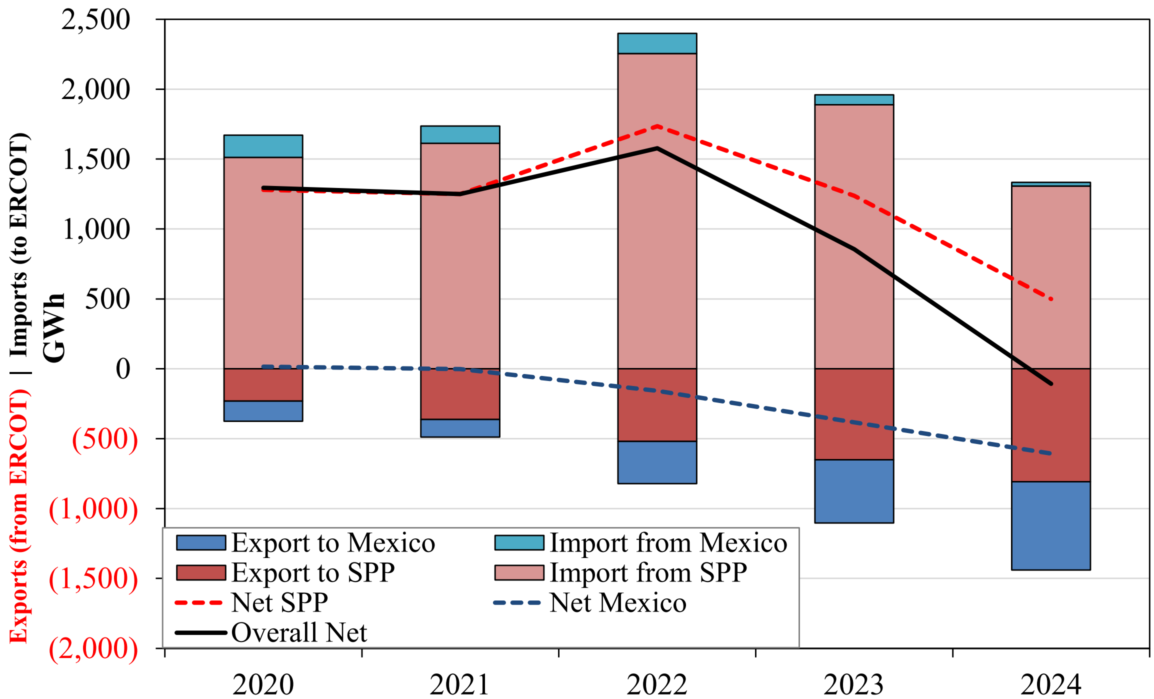
20 Capacity factor is the ratio of a resource's energy output to its maximum capability.

Figure 17 shows that in 2024, the peak net load was 69.6 GW, about 10 GW lower than in 2023 and 5 GW lower than in 2022. This decline was driven by a sharp increase in solar capacity and milder summer weather. However, 2024 is expected to be an exception. As the peak net load hour continues to shift later into the evening, when solar output falls off, peak net load is likely to rise rapidly in the coming years.

D. Imports to ERCOT

The ERCOT region is connected to neighboring systems through several direct current (DC) ties. Two ties with the Southwest Power Pool (SPP) provide 820 MW of transfer capability, while two others with Mexico's Comisión Federal de Electricidad with 400 MW of capability. Power can flow in either direction across these ties, with exports increasing demand and imports increasing supply in ERCOT. Figure 18 shows the total energy transacted over the DC ties annually since 2020.

Figure 18: Annual Energy Transacted Across DC Ties



In 2024, ERCOT was a net exporter for the first time since 2018. Although it remained a net importer from the SPP, net exports to Mexico were large enough to offset those imports. This shift was likely driven by lower average power prices in Texas compared to prior years.

E. Energy Storage Resources

Energy storage resources are a category of technologies designed to consume and store energy for later use. ESRs may include pumped hydro storage systems, compressed air energy storage (CAES), hydrogen and other power-to-gas systems, and a variety of other technologies. Because battery energy storage systems (BESSs) constitute virtually all ESRs in ERCOT, the term ESR is used interchangeably with batteries throughout this report. This section provides an overview of ESR fundamentals and summarizes the rapid influx of ESR capacity into ERCOT in recent years. We also discuss the evolution of ESR participation in ancillary service markets, their penetration into the markets for energy, and the revenue trends associated with this evolution.

1. ESR Fundamentals

ESRs have operating characteristics that distinguish them from traditional generation:

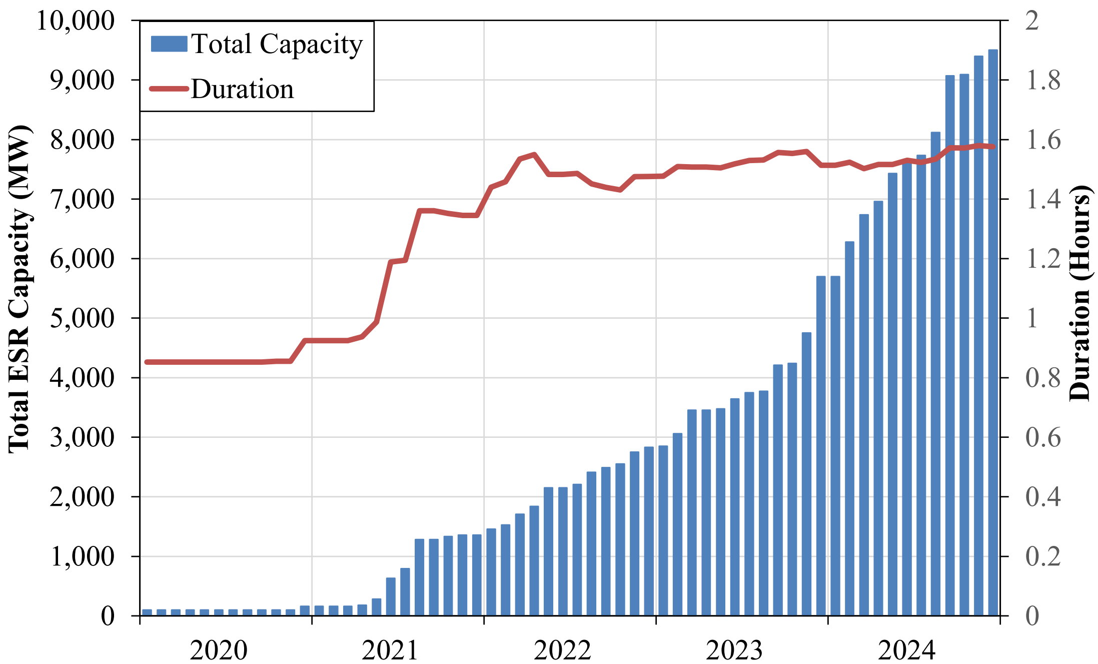
- **They are duration limited –** ESRs are limited by their state of charge (SOC), which represents the amount of energy they have stored at any given time. The average duration of batteries in ERCOT at the end of 2024 was 1.6 hours. This limitation has implications for how ESRs participate in the markets for energy and operating reserves.
- **They are a net load to the grid –** Even though ESRs function as both generation and load, they ultimately act as a net load on the system due to round-trip efficiency losses. ESRs must first consume energy to charge before later discharging, which is inherently inefficient with losses occurring through this cycle. The National Renewable Energy Lab (NREL) cites a roundtrip efficiency of 85% as its Annual Technology Baseline for 2024.[21]
- **Batteries ramp quickly –** Unlike thermal generators, which require time to start up and adjust their output, batteries are always "online" and can adjust their rate of charging and discharging almost instantly. This flexibility makes them well-suited for providing quick-responding ancillary services such as regulation or responsive reserves. Their ability to rapidly switch between charging and discharging makes them ideal for handling fluctuations in load or renewable generation.
- **They are driven by opportunity cost –** The charging/discharging cycle and the duration limitations create unique economics for ESRs. It requires ESRs to optimize their schedule for charging and discharging to maximize their revenue. Since the market software does not optimize intertemporally and only optimizes for one five-minute interval at a time, ESR owners must offer strategically to manage their SOC to maximize their value. These strategies can introduce market inefficiencies. A Multi-Interval Real-Time Market (MIRTM) would allow SCED to optimize charging and discharging of ESRs over time and reduce the need for ESRs to do this themselves. We recommended that ERCOT consider implementing a MIRTM in our 2022 State of the Market report.

[21] https://atb.nrel.gov/electricity/2024/utility-scale_battery_storage

2. ESR Capacity Trends

The first batteries started commercial operations in ERCOT in 2012. Their installed capacity remained below 300 MW until 2021 where it increased from 300 MW to 1,600 MW in only 12 months. ESR capacity has grown exponentially since then, reaching 9,505 MW by the end of 2024. Over that same time, the average duration of batteries has increased to approximately 1.6 hours. Figure 19 illustrates the growing trend of both the total installed capacity and average duration.

Figure 19: ESR Installed Capacity, 2020-2024

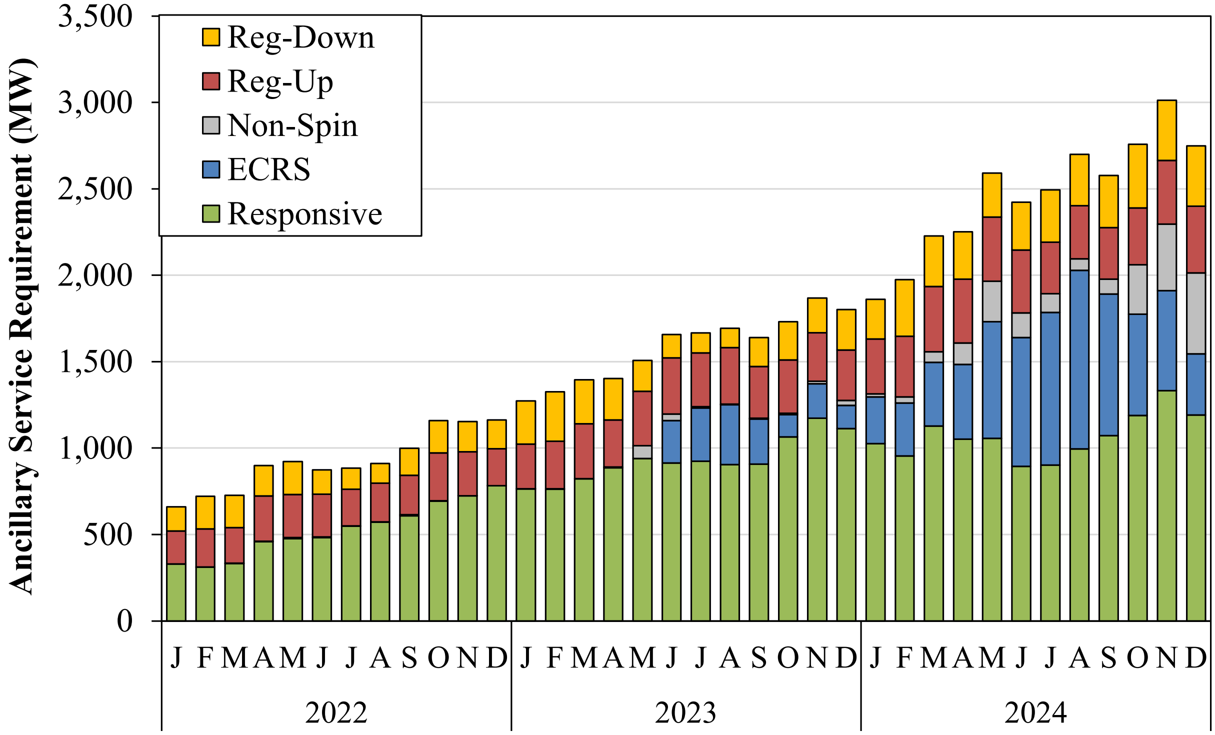


3. ESR Participation in Ancillary Service Markets

The rapid expansion in ESRs' provision of ancillary services like Regulation Up Reserve Service (Reg-Up), Regulation Down Reserve Service (Reg-Down), and Responsive Reserve Service (RRS) began even before 2024. In 2024, ESRs continued to increase their share of deployments across key ancillary services. On average, ESRs provided 84%, 77%, and 39% of Reg-Up, Reg-Down, and RRS, respectively. Their ability to ramp instantly and respond with precise, sub-second accuracy makes them ideal for frequency regulation, where continuous adjustments are needed to maintain system balance. Figure 20 illustrates the growing role of ESRs across ancillary service products.

It should be noted that batteries are able to provide two distinct types of responsive reserves. In addition to the primary frequency response product provided by conventional generators, batteries are also able to provide fast frequency response (FFR), a variation of RRS that operates via relay in response to frequency deviations. FFR was implemented in March 2020 through phase 1 of NPRR 863,[22] and the volume of RRS that can be provided by FFR is limited to 450 MW. We go into further detail on the limits imposed on FFR in the section on "Ancillary Services from Load Resources" later in this chapter.

Figure 20: ESR Participation in Ancillary Services, 2022-2024



ESRs also provide ERCOT Contingency Reserve Service (ECRS) and Non-Spin Reserve Service (NSRS), but their provision of these products is limited by the duration requirements for providing these products, two hours for ECRS and four hours for NSRS. These limitations disincentivize ESRs from providing these reserves in favor of participating in the market for energy. These incentives could be problematic under system-wide shortage conditions where it would be more promoting of reliability for duration limited resources to be held in reserve. Thus, we recommend that ERCOT reduce the duration requirement for both ECRS and NSRS to one hour in Recommendation 2024-2.

4. ESR Participation in Energy Markets

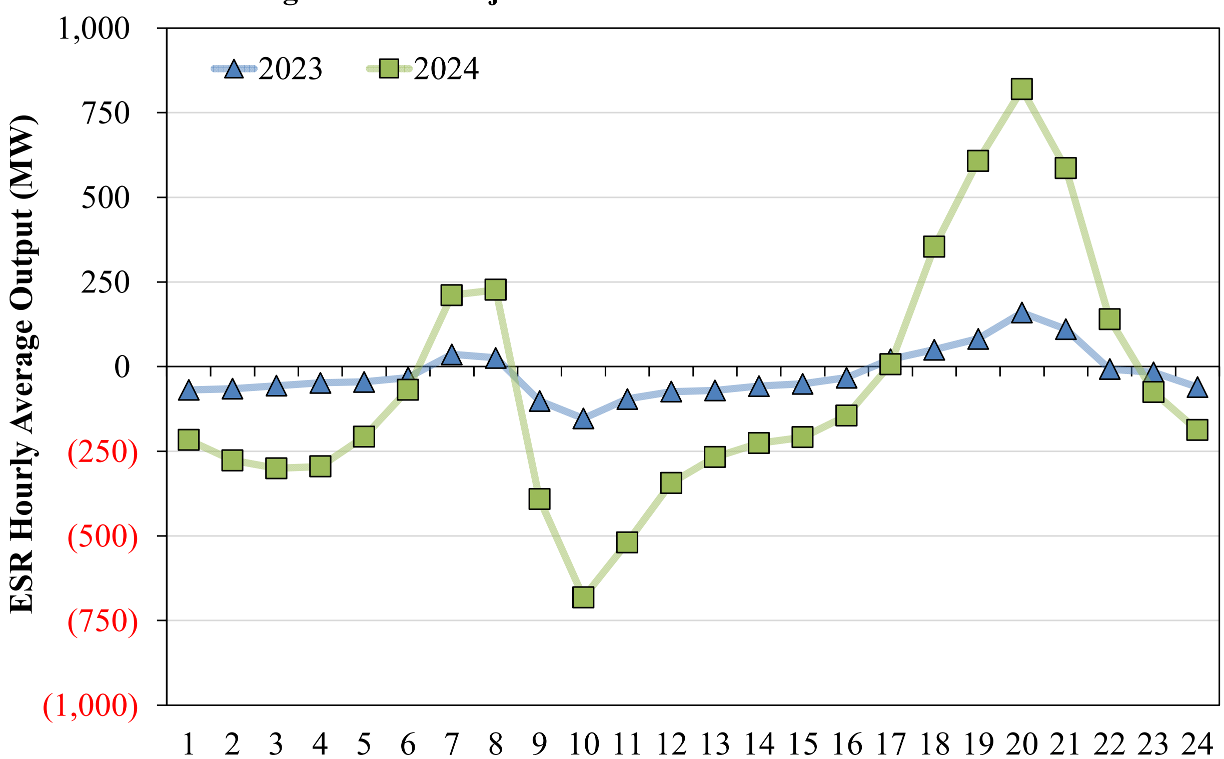
Prior to 2024, ESR participation in the wholesale market was primarily driven by provision of ancillary services, reflecting their ability to respond quickly to fluctuations in supply and demand. As ESRs saturated the market for several AS products, they shifted meaningfully toward energy arbitrage in 2024. Energy arbitrage refers to the practice of buying electricity when prices are low and selling it back to the grid when prices are high. It relies on the natural

22 https://www.ercot.com/mktrules/issues/NPRR863

fluctuations in electricity prices caused by supply and demand dynamics throughout the day, thus smoothing price volatility and enhancing grid stability.

The continued rapid influx of IRRs into the ERCOT market, particularly solar photovoltaic (PV), has also facilitated the growth of ESRs by enabling low-cost charging throughout the day. As solar generation has surged in the midday hours, real-time prices have dropped correspondingly, sometimes even to negative prices, allowing batteries to charge inexpensively or even get paid to consume energy that would otherwise be curtailed. Figure 21 illustrates this trend by showing net injections of energy from ESRs on an hour ending basis for 2023 and 2024. A positive injection value indicates aggregate discharging to the grid and a negative value represents aggregate charging. The figure highlights the dramatic increase in charging and discharging from ESRs in 2024 compared to the prior year.

Figure 21: Net Injection of Power from ESRs in 2024



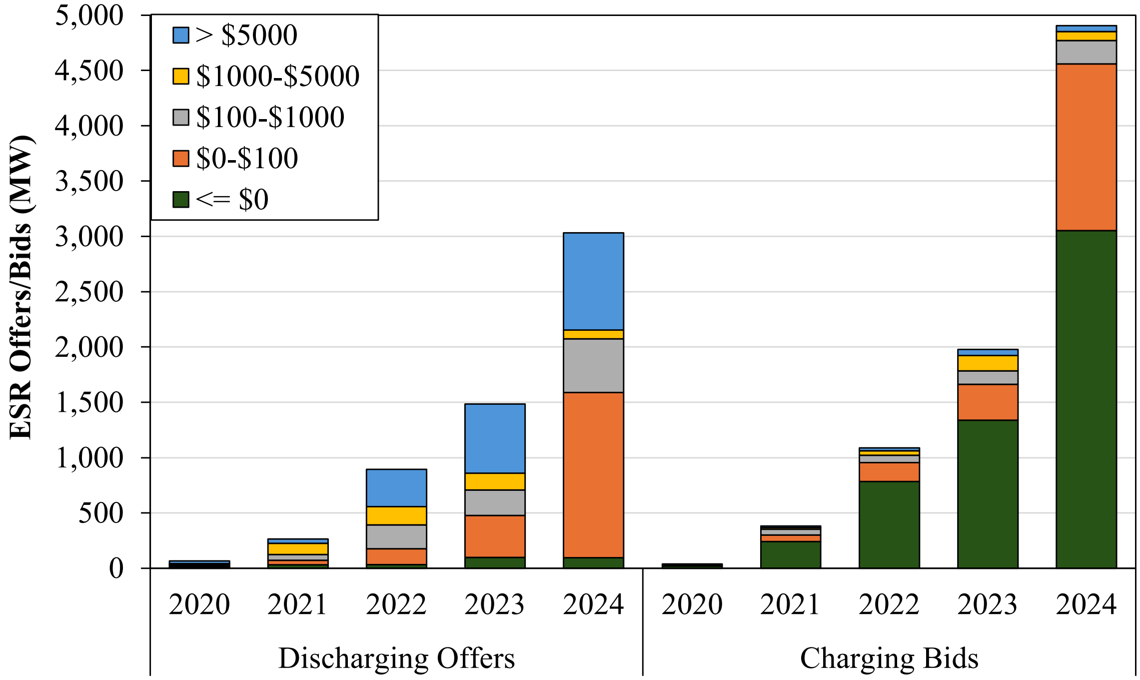
To further examine the evolution of ESR participation in the real-time energy market, we consider the aggregate bid and offer curves submitted by ESRs from 2020-2024, which is summarized in Figure 22. This figure shows a number of trends:

- The volume of energy offers submitted by ESRs has grown exponentially, doubling from almost 1,500 MW on average in 2023 to over 3,000 MW on average in 2024.
- The percentage of ESR offers that are priced competitively according to normal market clearing prices has also increased substantially, with nearly half of energy offers priced between $0-100 per MWh on average in 2024 compared to only 25% priced in that range

in 2023. This second trend suggests that as more ESRs have entered the market, they are forced to offer more competitively to maximize revenue from energy sales.

- ESRs still submit a substantial percentage of offers prices at the System-Wide Offer Cap (SWCAP), 29% in 2024 compared to 42% in 2023, which can partially be attributed to SOC management. ESRs must maintain sufficient SOC for satisfying their AS obligations, even if they are using some capacity for energy arbitrage. The drop from 2023 to 2024 is also reflective of the ECRS-induced artificial scarcity in 2023, which made it much more likely for offers priced at SWCAP to clear than in 2024.

Figure 22: Average Aggregate Offers for ESRs to Buy or Sell Energy, 2020-2024



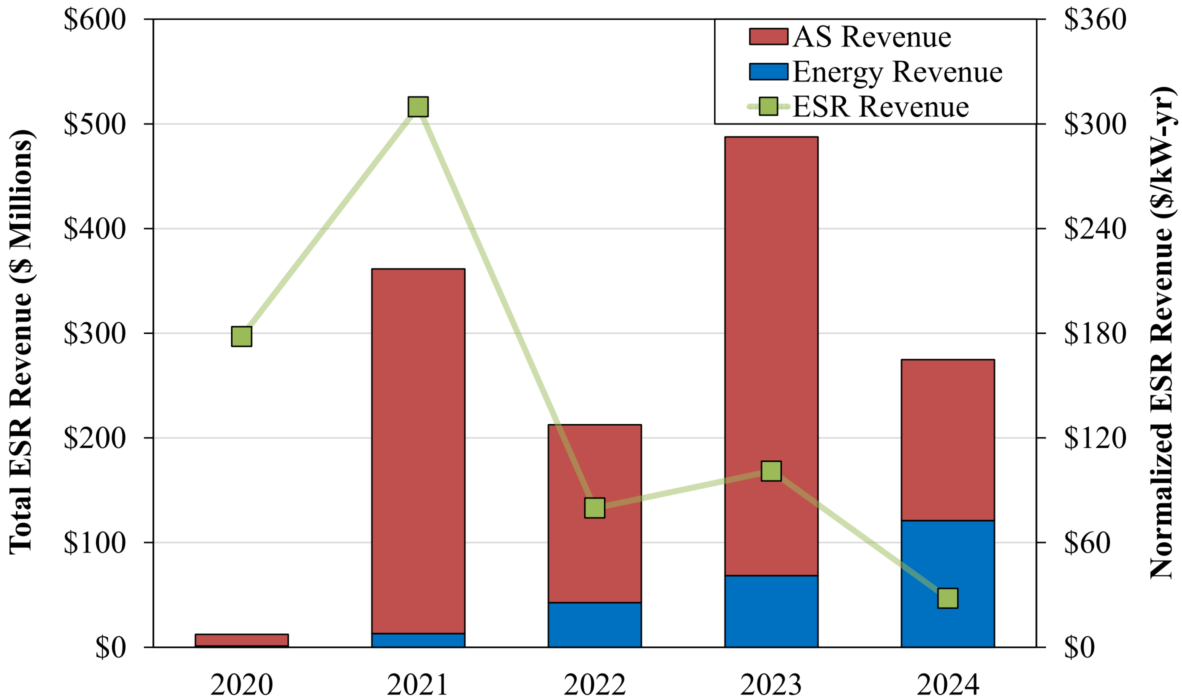
In contrast, Figure 22 also shows the aggregated bid curves for ESRs to purchase energy for charging. Since 2021, most charging bids have been priced at or below $0, reflecting the goal of minimizing charging costs to maximize net revenues from energy arbitrage. However, the share of charging bids priced above $0 has grown significantly in recent years. This suggests that ESRs are increasingly willing to pay for charging as competition in the energy market grows and opportunities for low- or zero-cost charging become less common.

As more ESRs enter the market, their collective charging demand will tend to push prices higher, reducing the frequency of $0 clearing intervals. It is also notable that the amount of ESR capacity offering to buy energy far exceeds the capacity offering to sell. This likely reflects that much of the ESR discharging capacity is committed to ancillary services, such as up-reserves, and is therefore not available to SCED.

5. ESR Revenue Trends

Economic fundamentals would suggest that as more ESRs enter the market, their revenues per unit of ESR capability will tend to decrease as the market for ancillary services becomes saturated and price spreads for energy arbitrage tighten. The revenue trends shown in Figure 23 validate these expectations, with some exceptions. Impactful events like Winter Storm Uri in 2021 and the implementation of ECRS in 2023 created revenue spikes that obscure a clear year-over-year trend in revenue per kilowatt (kW) or total revenue growth. In 2024, total ESR revenue declined compared to 2023, primarily because of the increasing supply surplus from new IRRs and fewer extreme heat days that together reduced the frequency of price spikes. Normalized revenues from energy arbitrage have grown every year as their offer patterns have become more competitive.

Figure 23: Total and Normalized ESR Revenue, 2020-2024



F. Demand Response Capability

Demand response is a term that refers to actions that can be taken by consumers of electricity to reduce their load in response to instructions from ERCOT or economic incentives. Examples of demand response in ERCOT include operating reserves provided by load resources, price-responsive dispatch through SCED, self-curtailment based on economic incentives, and reliability programs administered by ERCOT and transmission and distribution service providers (TDSPs). Trends related to each of these forms of demand response are discussed in this section, but first we examine the proliferation of Large Flexible Loads in the ERCOT market in recent years.

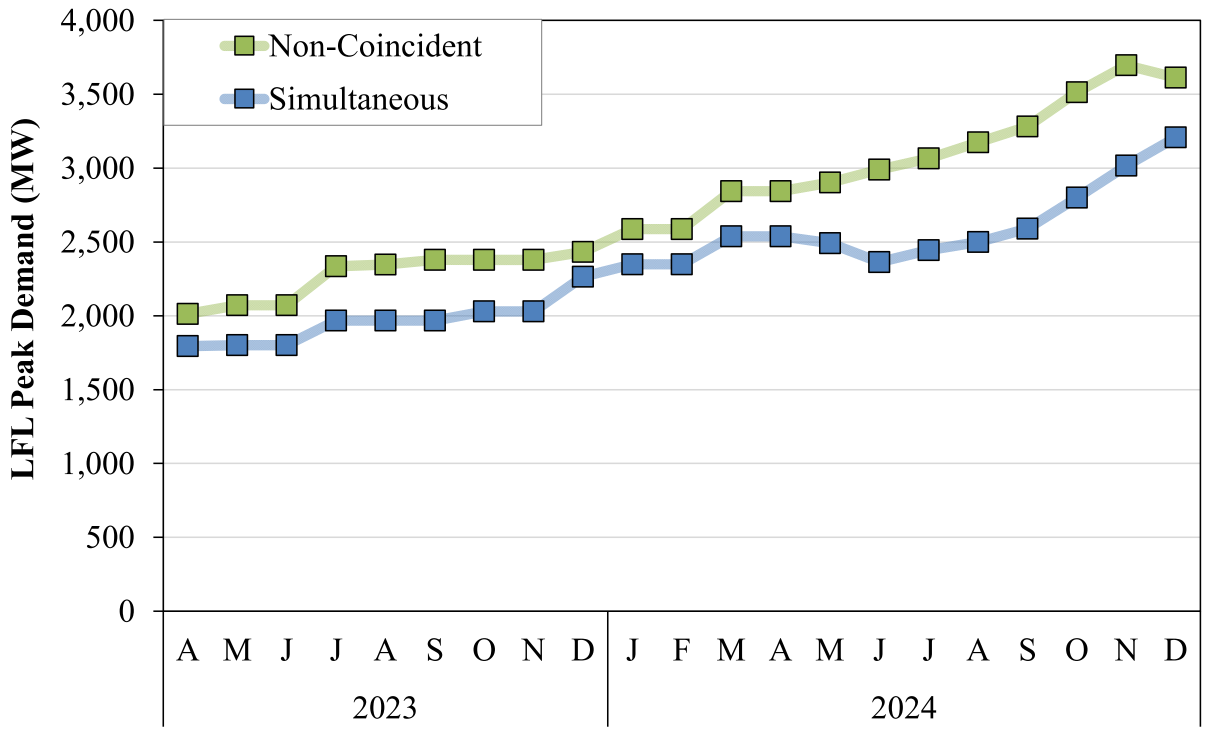
1. Large Flexible Loads

Large Flexible Loads (LFLs) are characterized by their significant size, typically ranging from tens to hundreds of MWs per interconnection, and their ability to adjust consumption based on market conditions. They routinely participate in demand response to optimize the economics of their operations. The two primary types of demand response used by LFLs are price response, which involves reducing consumption to avoid high energy prices, and Four Coincident Peak (4CP) response, which involves reducing consumption during system peak intervals to manage transmission cost exposure. These forms of demand response are discussed in more detail in the following subsections.

The vast majority of LFLs in ERCOT are cryptocurrency mines. These operations consist of highly specialized data centers built to perform the complex computations required to validate cryptocurrency transactions. Unlike traditional data centers, cryptocurrency operations can rapidly adjust their load, and their profitability is very sensitive to the price of electricity, so they will generally have incentives to reduce consumption when prices are high. This makes them well suited to participate in price-based demand response.

Crypto-mining operations began relocating to Texas in large numbers in 2021 following a nationwide ban in China. By spring 2023, more than 2,000 MW of LFL capacity had already been identified. By the end of 2024, that number had grown to over 3,600 MW of non-coincident demand and more than 3,200 MW of simultaneous demand, as shown in Figure 24.

Figure 24: Non-Coincident and Simultaneous Peak Demand for LFLs



Some notable characteristics of LFLs in ERCOT as of the end of 2024 include the following:

- Approximately 30% of LFLs are co-located with generation.
- More than half of LFL capacity is located in the West load zone, followed by the North and South zones.
- A substantial percentage of LFL capacity has been registered as Load Resources, though the exact percentage has varied as LFLs have changed ownership and resource status.

The next subsections focus specifically on load resources, first their role in providing ancillary services, and then their participation in the real-time market as price-responsive load.

2. Ancillary Services from Load Resources

ERCOT allows qualified load resources to offer into the day-ahead ancillary services markets. There are two types of load resources in ERCOT – non-controllable load resources (NCLRs) and controllable load resources (CLRs). Because of their different capabilities in responding to system conditions or instructions from ERCOT, CLRs and NCLRs have different limitations in the types and volumes of ancillary services they can provide.

Ancillary Services from NCLRs

As of November 2024, there were 505 registered NCLRs in ERCOT, with a combined capacity of 10,016 MW.[23] NCLRs can qualify to provide RRS, NSRS, and ECRS. Historically, their primary contribution has been to provide responsive reserves. Table 3 shows the average ancillary services supplied by NCLRs from 2022 through 2024.

Table 3: Average Volume (MW) of Ancillary Services Provided by NCLRs, 2022-2024

	RRS	ECRS	NSRS
2022	1,769	N/A	10
2023	1,742	67	70
2024	1,700	164	3

Although more than 8,300 MW of NCLR capacity was qualified to provide responsive reserves in 2024, the average amount actually provided was only 1,700 MW. One reason for this gap is that ERCOT imposes limits on the share of total responsive reserves that can come from NCLRs. Conventional generators provide responsive reserves through primary frequency response, which automatically adjusts output in response to large frequency deviations. In contrast, NCLRs provide responsive reserves either by automatically disconnecting through an under-frequency

23 *See* 2024 Annual Report of Demand Response in the ERCOT Region (Feb. 2025), available at https://www.ercot.com/mp/data-products/data-product-details?id=NP3-110

relay (UFR) at 59.7 Hertz (Hz) or by responding to manual deployment instructions from ERCOT during an Energy Emergency Alert Level 2.

ESRs, which as noted earlier can provide FFR, represent a third method of delivering responsive reserves. ERCOT limits the combined volume of responsive reserves from UFR and FFR to no more than 60% of the RRS volume in the Ancillary Services Plan for each hour. In addition, FFR is subject to a separate cap of 450 MW. Rather than clearing each type of responsive reserve separately with distinct prices, ERCOT clears them together and prorates the awards for UFR and FFR when the combined limit is binding. Figure 25 shows this by comparing the volumes of UFR and FFR that are self-scheduled or awarded in the day-ahead market (DAM) and the remaining uncleared offers and the ERCOT limit.

Figure 25: Responsive Reserves from UFR and FFR in DAM

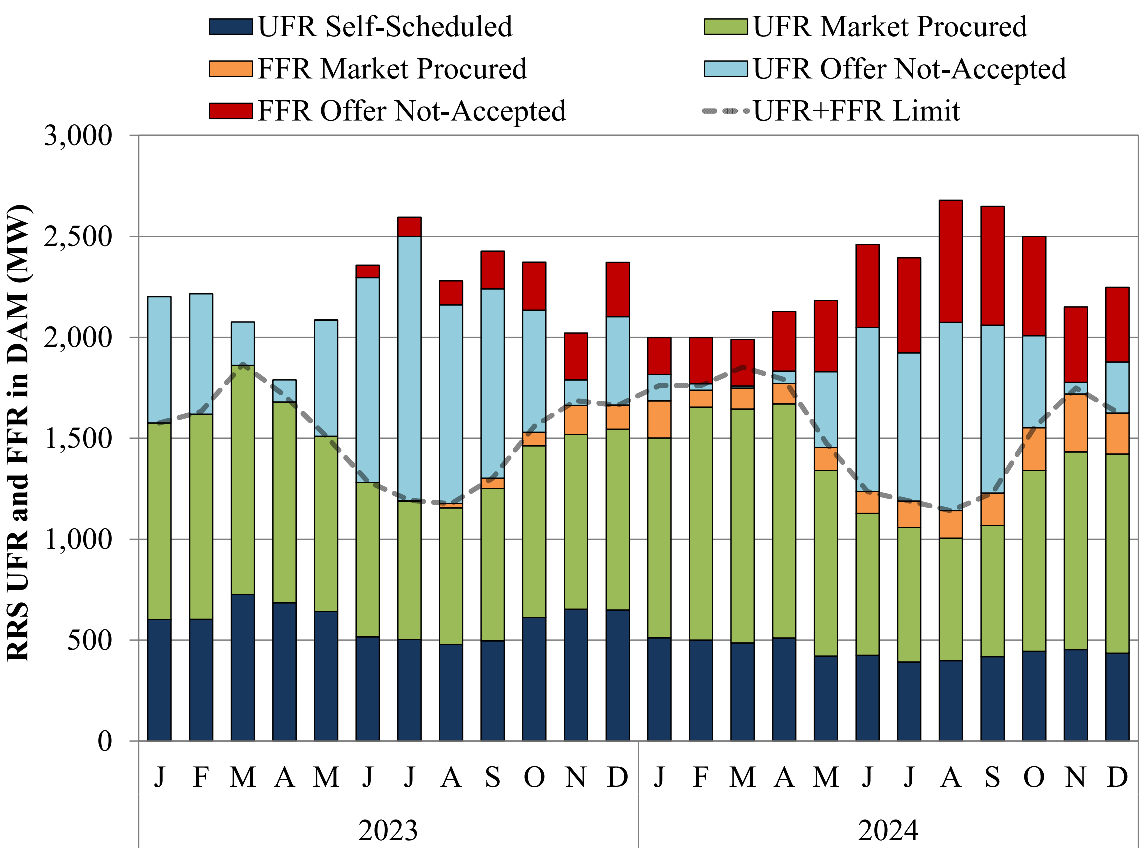


Figure 25 shows that the offers to provide UFR decreased from 2023 to 2024, likely because of the decrease in RRS prices over this timeframe made it less economic for some of the NCLRs to be willing to curtail.

Interestingly, the limit on UFR plus FFR is no longer consistently binding, as primary frequency response (PFR) offers from ESRs are often more competitive. In addition, total FFR offer volumes are frequently below the 450 MW cap. This trend reflects a shift in ESR offer strategy. Since providing PFR is similarly costly, many ESRs prefer to offer PFR to avoid the risk of their

FFR offers being prorated when the combined limit is reached. This behavior is likely a direct result of ERCOT clearing multiple forms of responsive reserves together, rather than treating them as distinct products with separate clearing prices, as we address in Recommendation 2019-2.

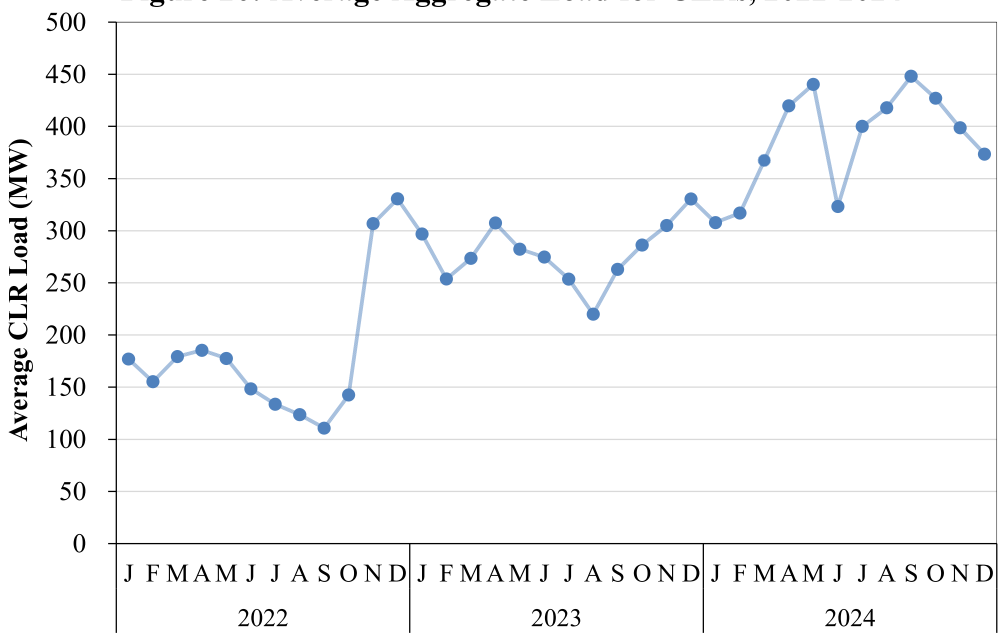
In addition to providing RRS, load resources can also participate in NSRS and ECRS. As with RRS, ERCOT places a cap on how much ECRS can come from NCLRs, limiting their share to no more than 50% of the total Ancillary Services Plan for ECRS. As shown in Table 3, NCLR participation in ECRS remains low, despite ECRS consistently clearing at a premium price to RRS and carrying less risk of proration at current volumes. The reasons for this low participation are unclear.

NCLRs first began participating in NSRS in November 2022 following the implementation of NPRR 1093[24] and NPRR 1101.[25] However, their participation has remained limited. To qualify to provide NSRS, NCLRs cannot have an active UFR, which means they must choose between providing NSRS or RRS. Since RRS is generally the more valuable product, most NCLRs opt to participate in RRS instead.

Ancillary Services from CLRs

Load telemetry from CLRs averaged only 300-450 MW of across all of 2024 as shown in Figure 26.[26] Much of this load can be attributed to CLRs carrying ancillary services.

Figure 26: Average Aggregate Load for CLRs, 2022-2024



[24] https://www.ercot.com/mktrules/issues/NPRR1093

[25] https://www.ercot.com/mktrules/issues/NPRR1101

[26] The set of CLRs included in this data provided by ERCOT's demand integration team does not include ESRs

CLRs that are capable of providing PFR are eligible to provide regulation, RRS, NSRS, and ECRS. The amount of responsive reserves they can offer depends on their droop setting, and there is no cap on the share of responsive reserves in the Ancillary Services Plan that can come from CLRs. NPRR 1244, approved in November 2024,[27] expanded eligibility for CLRs that cannot provide PFR, allowing them to participate in NSRS and ECRS while disqualifying them from providing regulation or responsive reserves.

Table 4 shows the average volume of operating reserves provided by CLRs from 2022 through 2024.

Table 4: Average Volume of Ancillary Services Provided by CLRs, 2022-2024

	RRS	ECRS	NSRS
		MW	
2022	82	N/A	0
2023	36	58	9
2024	34	37	47
8/23/24-12/31/24	45	22	82

The volume of regulation provided by CLRs is effectively zero, because regulation service carries a much higher opportunity cost compared to other ancillary services. Regulation is deployed frequently, which reduces a facility's available run time and limits revenues from its primary operations. As shown in Table 4, CLRs primarily provided RRS in 2022. In 2023, following the introduction of ECRS, CLR participation shifted toward ECRS, which offered higher compensation than RRS.

Until the implementation of NPRR 1131 in August 2024, CLRs could only provide offline NSRS, which required them to continue consuming electricity until receiving a manual instruction from ERCOT to reduce load. This exposed participants to potentially high real-time prices. Since NPRR 1131 took effect, CLRs have been allowed to provide online NSRS, with their capacity treated as a standing deployment that SCED can dispatch down based on an energy bid curve priced at no less than $75 per MWh. As a result, the volume of NSRS provided by CLRs increased substantially after August 2024.

Relative to their total capacity, the average volume of ancillary services provided by CLRs remains small. One reason for this low participation is that, under the current market design, energy and ancillary services are not co-optimized for load resources in the day-ahead market. As a result, if a load resource is awarded ancillary services in DAM, it remains exposed to real-time energy prices. These incentives are expected to improve significantly with the implementation of Real-Time Co-optimization (RTC) and, more notably, with NPRR 1188,

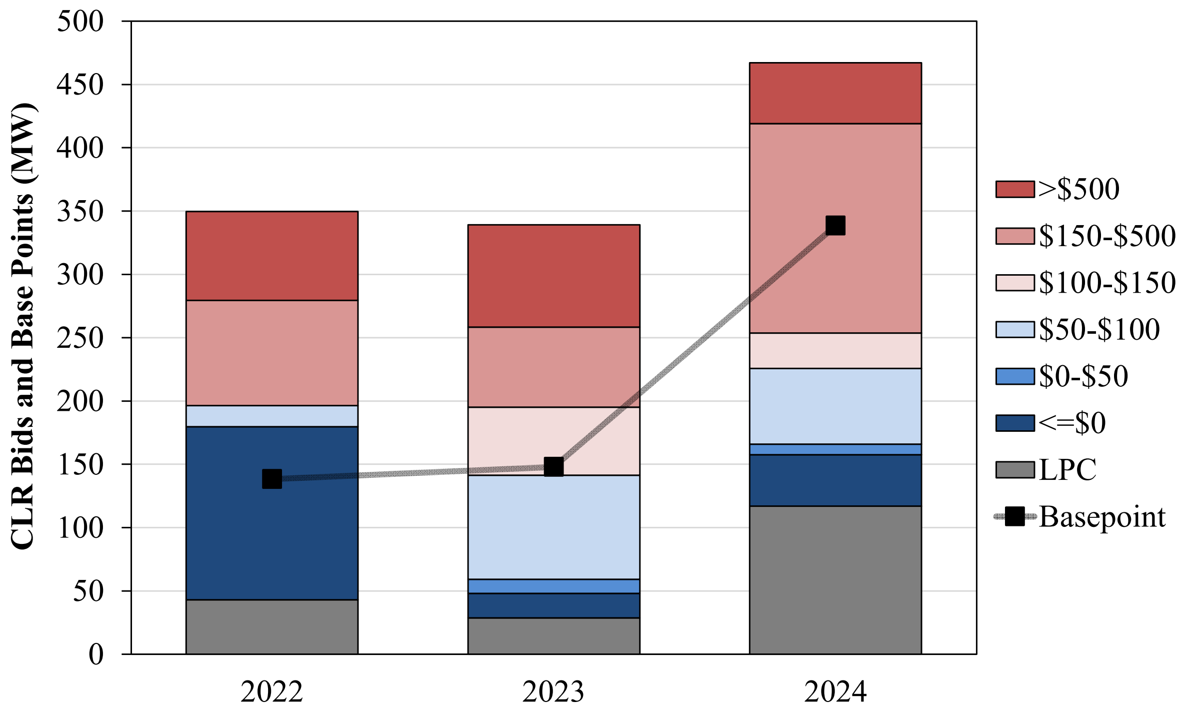
27 https://www.ercot.com/mktrules/issues/NPRR1244

which will introduce day-ahead co-optimization of energy and ancillary services for load resources.

3. Price Response through SCED

The participation of CLRs in SCED was implemented in 2014, allowing loads that can respond to 5-minute dispatch instructions to submit bids to buy electricity at certain price quantity pairs, which allows them to be dispatched down when the clearing price exceeds these bids. The average aggregate bid prices and corresponding base points associated with CLRs for 2022-2024 are shown in Figure 27.

Figure 27: Average Aggregate Bid Curves and Base Points for CLRs, 2022-2024



Note that LPC refers to the aggregate low power consumption level of CLRs, or the minimum level to which they can be dispatched down by SCED. Year-to-year variation in CLR bid prices reflects the changing economic value of their output, including fluctuations in cryptocurrency prices and the associated break-even cost of electricity consumption. For example, the share of CLR bids priced above $100 increased from 50% in 2022 to 69% in 2024. This shift reflects both rising cryptocurrency prices and improvements in mining hardware efficiency. As a result, CLRs received a greater volume of base points, since they were dispatched down less frequently due to prices clearing above their submitted bids.

Several upcoming market design changes are expected to improve the incentives for loads to register as CLRs. One of the most important is the implementation of RTC, which will reduce the risk that loads face from high real-time prices when fulfilling ancillary service obligations.

Another major change will come through NPRR 1188, approved by the PUCT in November 2024, which will introduce nodal pricing for CLRs.

Currently, CLRs are dispatched and settled using zonal prices, which creates two economic challenges. First, for CLRs located at nodes with structurally low energy costs, such as those in the Panhandle near large volumes of renewable generation, the zonal price is often higher than the cost of serving their load. Second, CLRs are still settled at the zonal price even when dispatched down in response to price spikes. For example, if a CLR has a break-even price of $100, and the zonal price across three SCED intervals is $50, $50, and $350, the average settlement price over those 15 minutes would be $150. Although the CLR would be dispatched down during the $350 SCED interval, it would still incur a loss over the full settlement interval.

In addition, NPRR 1244, also approved in November 2024, will allow loads to register as CLRs even if they cannot provide PFR. This expands participation to wholesale consumers that can be dispatched in SCED, improving both reliability and price formation in real time, even if those loads cannot provide the full suite of ancillary services. Taken together, these market design improvements could lead to a significant increase in the number of loads registering as CLRs.

4. Self-curtailment

In addition to participating in demand response programs that involve direct instructions from ERCOT to reduce consumption, loads also engage in self-curtailment, which occurs outside of the formal wholesale market process. There are two primary forms of self-curtailment. The first is price response, where a load voluntarily reduces consumption to avoid high prices without receiving a dispatch instruction from SCED. The second is 4CP response, where a load reduces consumption during system peaks to lower its share of transmission cost allocation.

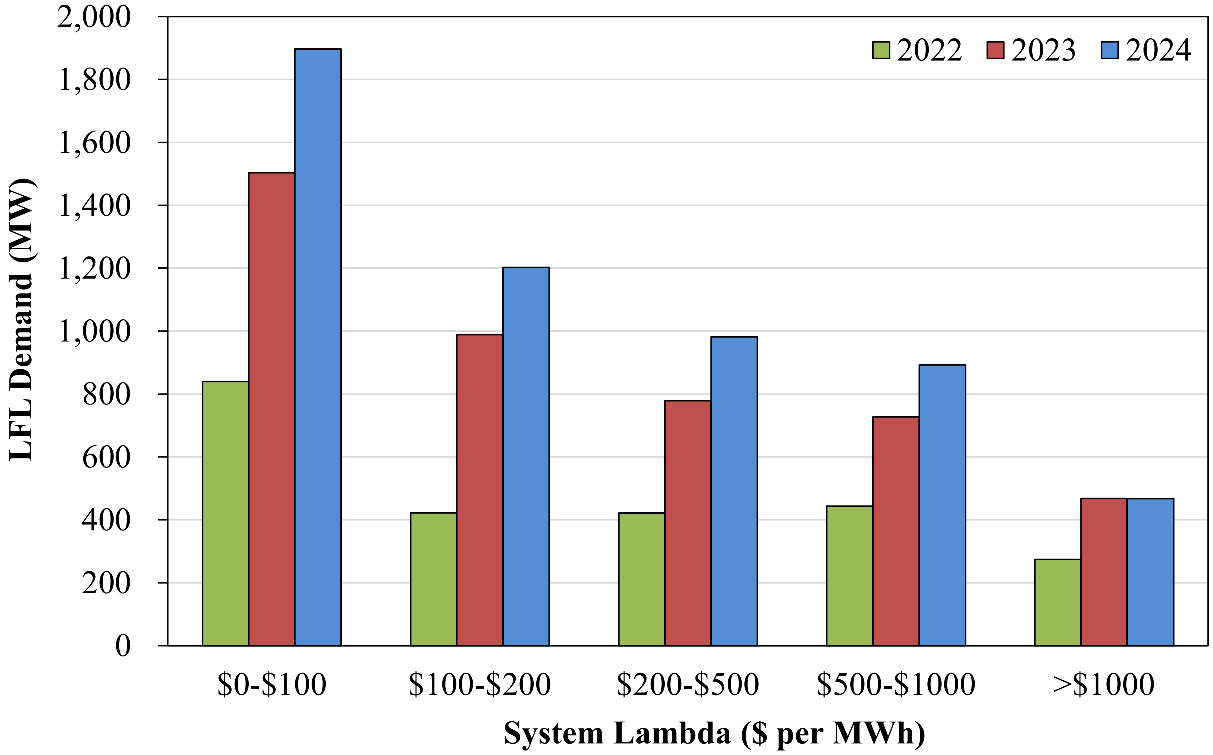
Self-curtailment through price response has grown significantly with the rise of demand from LFLs. To illustrate this trend, Figure 28 shows the aggregate load from LFLs[28] as a function of system-wide prices. The data indicate that LFLs reduce their consumption by as much as 75% when system lambda exceeds $1,000 per MWh.

Self-curtailment as price response from LFLs is a relatively recent trend. Historically, the more consequential form of self-curtailment has been 4CP response. Under this mechanism, transmission costs are allocated to load serving entities based on their load ratio share during the highest 15-minute system load intervals in each of the four months from June through September, a method known as 4CP. By reducing demand during these intervals, entities can

28 Note that these data do not include consumption from CLRs because they are dispatched by SCED rather than self-curtailed

lower their share of transmission charges, which were approximately $10 per MWh in 2024.[29] ERCOT estimates that about 2,500 MW of load was reduced during 4CP intervals in 2024, roughly 1,000 MW less than in 2023.[30] This year-over-year decline may be attributed to reduced overall load due to milder summer weather and fewer extreme events.

Figure 28: LFL Demand vs. System Lambda, 2022-2024



While both forms of self-curtailment can help reduce system stress during periods of high demand, they also present challenges for grid operations and market design. For price-based curtailment outside of SCED, sudden drops in load are not captured in economic dispatch and must be balanced using operating reserves. As the volume of this type of response has increased, so has the need for additional reserves to maintain real-time balance between supply and demand. To address this issue, ERCOT should encourage more loads to register as CLRs, allowing them to be dispatched down through SCED rather than reacting independently. The implementation of RTC and nodal pricing for CLRs through NPRR 1188 should significantly improve the incentive structure for this transition.

29 ERCOT Report on Existing and Potential Electric System Constraints and Needs, December 2024, https://www.ercot.com/files/docs/2024/12/20/2024-report-on-existing-and-potential-electric-system-constraints-and-needs.pdf

30 *See* ERCOT, 2023 Annual Report of Demand Response in the ERCOT Region (Jan. 2024) at 18, available at http://www.ercot.com/services/programs/load.

In the case of 4CP-related self-curtailment, several adverse effects have emerged. Although transmission costs were once closely tied to peak demand, this relationship is weakening. Load growth in regions like West Texas, away from traditional population centers, is driving transmission needs regardless of system peak. Additionally, with the increased role of renewables, peak demand is no longer a reliable proxy for scarcity or high prices. As a result, load reductions during peak intervals can suppress prices and reduce revenues to generators, which in turn poses long-term risks to resource adequacy. To address these distortions, we continue to recommend revisiting the transmission cost allocation methodology in favor of a framework based on cost causation principles (see SOM Recommendation 2015-1).

5. Reliability Programs

There are two main reliability programs in which ERCOT loads can participate. The first is the Emergency Response Service (ERS) program, administered by ERCOT. The second consists of demand response programs managed by the transmission and distribution service providers (TDSPs). The ERS program was established by a PUCT rule adopted in March 2012 and has a program budget of $75 million as of August 2022.

Industrial and commercial electricity consumers submit offers to provide ERS through a centralized auction, and ERCOT deploys ERS during Energy Emergency Alerts (EEAs) or when Physical Responsive Capability (PRC) falls below 3,000 MW. Since its inception, ERS has been deployed for nine events, most recently in September 2023. For the 2024 program year (December 2023 through November 2024), ERCOT procured an average of 1,147 MW per hour, an 18% increase from 2023, at an average clearing price of $7.36 per MWh, a 5% decrease from the prior year.

TDSP Load Management programs operate similarly by procuring demand response capacity from end-use customers, but only for the summer and winter peak seasons. While TDSPs generally manage the testing and deployment of their enrolled load resources, those resources may also be dispatched by ERCOT during an EEA Level 2 event. In 2024, approximately 113 MW of load participated in the winter program and 289 MW participated in the summer program, reflecting year-over-year increases of 22% and 11%, respectively.

III. DAY-AHEAD MARKET PERFORMANCE

ERCOT's day-ahead market (DAM) allows participants to take financially binding forward positions as hedges against real-time market outcomes. Examples of the types of positions market participants can take in the day-ahead market include the following:

- Purchases and sales of energy for delivery in real-time
- Hedges against the incremental value of congestion between day-ahead and real-time, i.e., point-to-point (PTP) obligations and options
- Sales of ancillary services to be provided in real-time

The DAM plays a critical role in coordinating generator commitment decisions and helping participants manage or arbitrage real-time price exposure. Although ancillary services involve physical delivery, the DAM is otherwise a voluntary financial market and creates no physical obligations from market awards. All bids and offers are cleared respecting transmission network constraints, producing nodal prices that reflect modeled congestion. In addition, the DAM provides an opportunity to hedge exposure to real-time price.

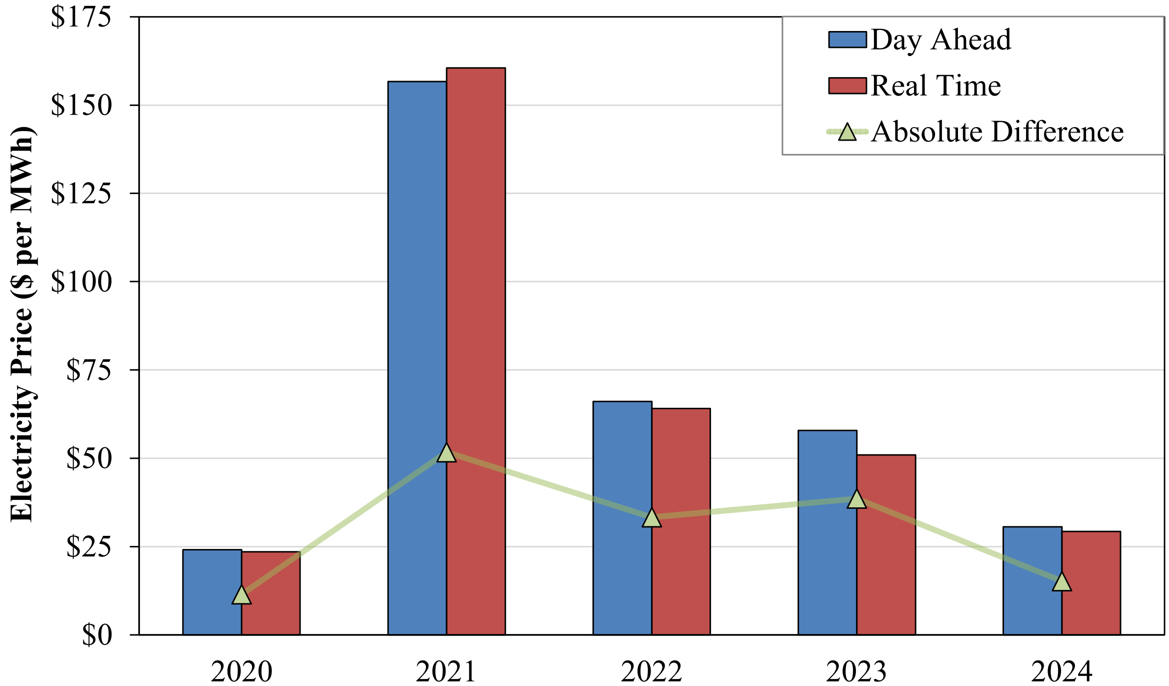
In this section, we examine day-ahead energy prices in 2024 and their convergence with real-time prices. We also review the activity in the DAM, including physical and virtually scheduled sales of energy and a discussion of PTP obligations. The section concludes with a review of the day-ahead ancillary service markets.

A. Day-Ahead Energy Market Pricing

A primary indicator of forward market performance is the extent to which forward prices converge with real-time prices over time. Prices should converge when: (1) there are low barriers to purchases and sales in either market, (2) sufficient information is available to allow market participants to develop accurate expectations of the real-time prices, and (3) the physical limitations of the transmission network are accurately reflected in both markets. These conditions allow market participants to arbitrage predictable differences between day-ahead and real-time prices, ultimately resulting in price convergence. Price convergence between the day-ahead and real-time markets is important because it leads to more efficient commitment of resources to be used in real-time.

The average price difference between day-ahead prices and real-time spot prices reveals whether persistent and predictable differences exist that participants should arbitrage over the long term. Figure 29 shows the annual average day-ahead and real-time prices for the past five years. It also shows the average of the absolute value of the difference between the daily average day-ahead and real-time price. This measure captures the volatility of the daily price differences, which may be large even if the prices converge on average.

Figure 29: Convergence Between Day-Ahead and Real-Time Energy Prices



Price convergence was much better in 2024 compared to recent years. The average price for energy in the day-ahead market was less than $1.50 per MWh higher than in the real-time market and the average absolute difference in prices was only $17.35 per MWh, the lowest since 2020.

B. Day-Ahead Market Activity

Market participants can participate in the day-ahead market by submitting bids to buy electricity or offers to sell electricity for real-time delivery. These bids and offers may take two forms:

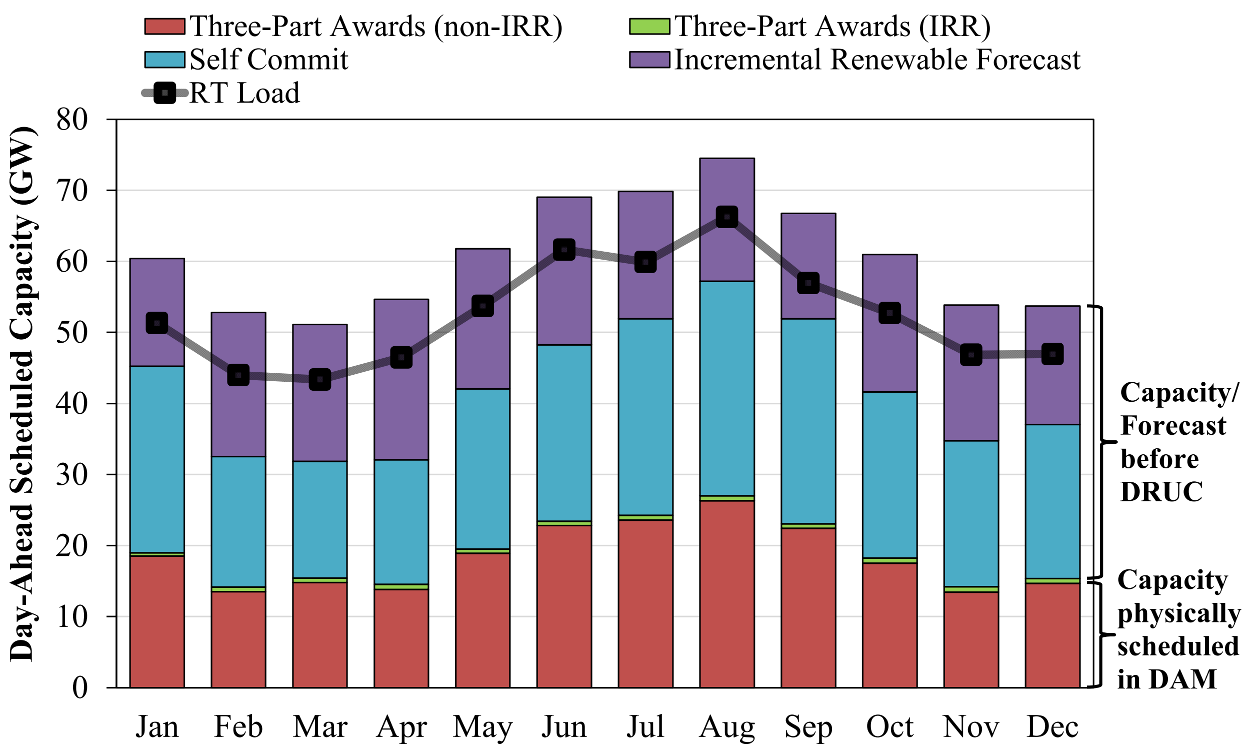
- A three-part supply offer, which allows a seller to represent the financial and operational details of a specific generation resource. This includes the startup cost, minimum generation cost, and an incremental energy offer curve.
- An energy-only bid or offer, which is a location-specific transaction not tied to an actual generation resource or load. These are referred to as virtual bids and offers, submitted either as a hedge or for speculative purposes.

1. Day Ahead Market Volume

The day-ahead market clears offers and bids by matching supply and demand. In 2024, the volume of day-ahead energy purchased through generator-specific offers and virtual energy offers equaled 61% of real-time load, in line with recent years. Participants also use PTP obligations scheduled in the day-ahead market to hedge congestion and other transactions outside the ERCOT wholesale market to hedge exposure to real-time prices.

Less than half of the resources are scheduled through the day-ahead market so the balance of the supply needed to satisfy real-time load must be provided by resources scheduled after the day-ahead market, as shown in Figure 30. Figure 31 shows awards for three-part offers (i.e., three-part awards) for thermal and renewable resources, thermal capacity self-scheduled in Current Operating Plans (COPs) prior to the day-ahead RUC process (DRUC), and the day-ahead forecast for renewable generation.

Figure 30: Day-Ahead Schedule/Forecast of Physical Generation



While the volume of generation cleared through three-part offers is limited, self-scheduled thermal resources and expected renewable output regularly supplement these awards. As a result, ERCOT does not typically need to rely on RUC to serve the majority of real-time load.

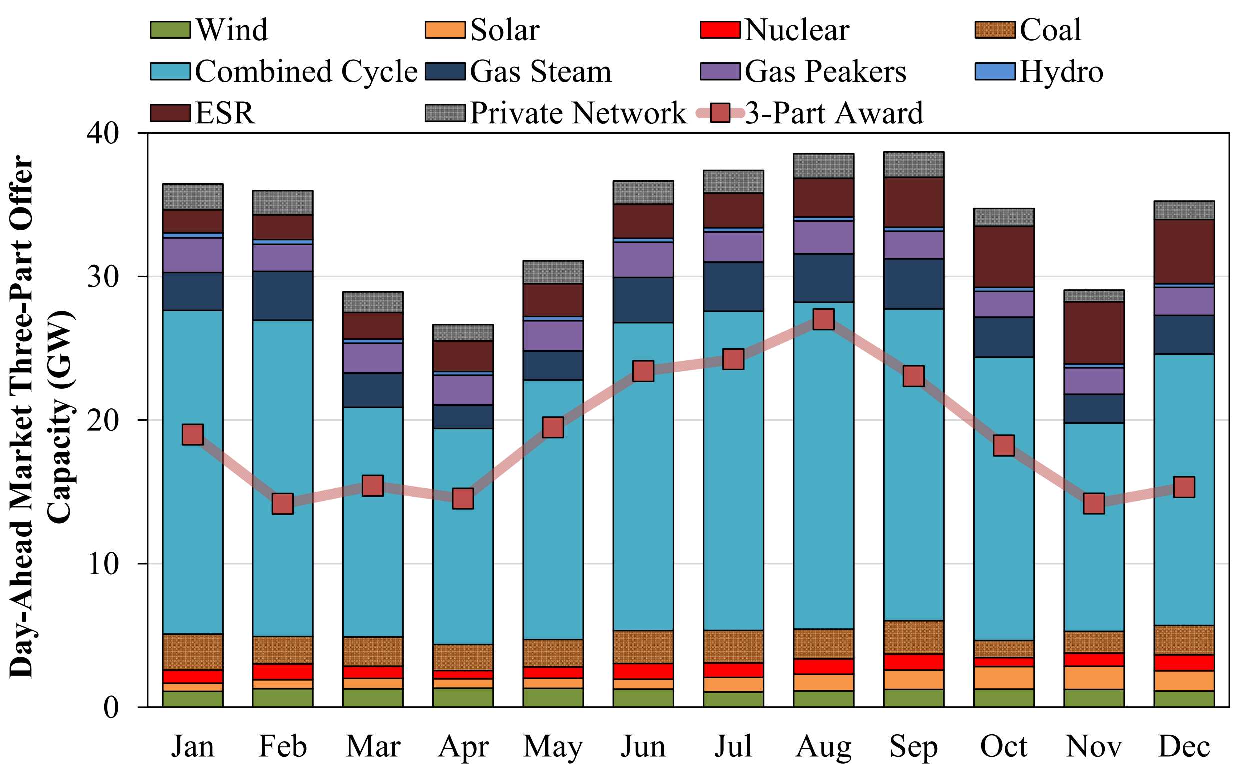
2. Three-Part-Offer Behavior

A persistent trend in the day-ahead market is the limited role of physical generation scheduled through awards to three-part offers. These awards account for just 59% of energy sold in the day-ahead market and serve only 36% of real-time load. This raises two key questions: (1) how frequently are qualified scheduling entities (QSEs) submitting three-part offers in the day-ahead market, and (2) to what extent are resources self-scheduling and accepting real-time price exposure.

The first question can be answered by looking at the volume and clearance rates of three-part offers submitted by QSEs. Figure 31 presents the average monthly capacity of three-part offers in 2024, broken out by fuel type. QSEs regularly submitted a substantial capacity of three-part offers, and on average, 56% of that capacity cleared in the day-ahead market. Clearance rates increased during the summer, when higher demand made thermal resources more economically competitive. Combined cycle natural gas plants accounted for the majority of three-part offers, indicating active participation from dispatchable resources.

The second question is reflected in the relatively low share of real-time load served by three-part awards. With only 36% of real-time load covered by these awards, many resources are self-scheduling and exposing themselves to real-time prices. For thermal resources, this exposure to real-time prices means that their start-up and minimum generation costs might not be covered by market revenues. Renewables resources, however, tend to avoid three-part offers because of uncertainty in their output, which introduces the risk of imbalance payments if their real-time generation is less than what they sold in the day-ahead market.

Figure 31: Day-Ahead Market Three-Part Offer Capacity



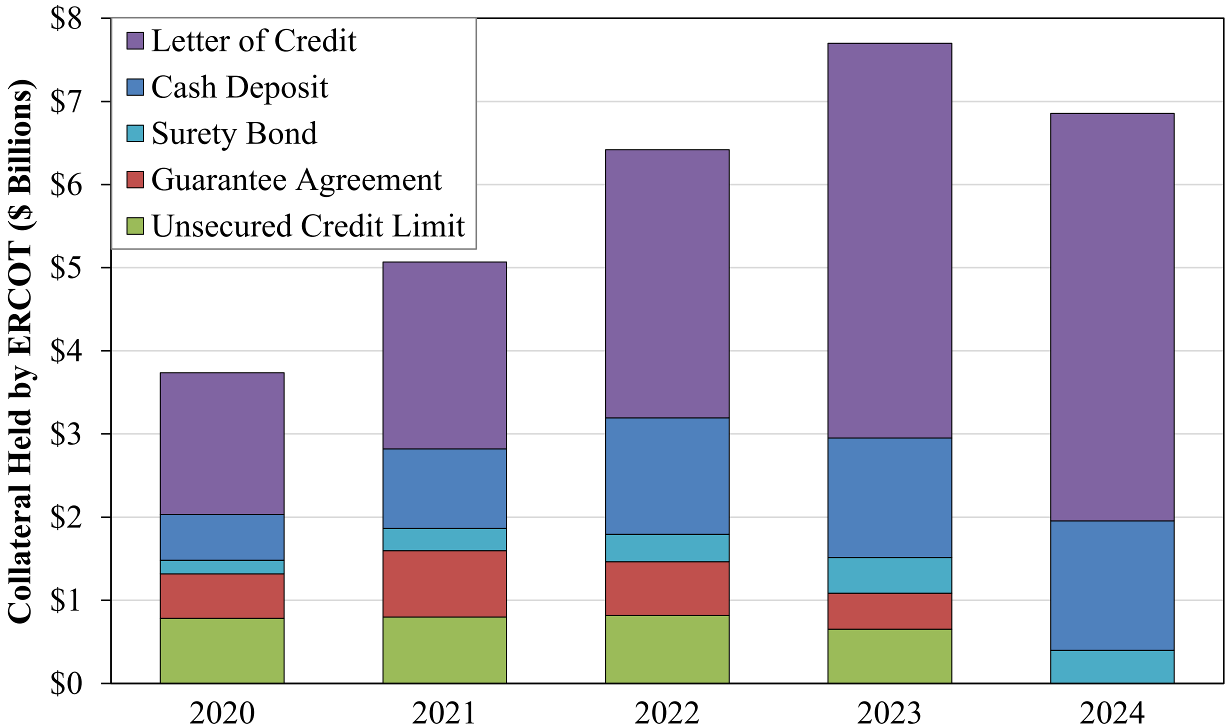
3. Collateral Requirements

The relatively low level of capacity submitting three-part offers suggests that there may be a disincentive to participate in the day-ahead market. One plausible explanation is that the collateral requirements to participate in the day-ahead market may be higher than QSEs think is worthwhile. To participate in ERCOT's day-ahead market, a market participant must have sufficient collateral with ERCOT, and these credit requirements are a constraint on submitting bids in the day-ahead market. When the available credit of a QSE is limited or expensive, its participation in the day-ahead market will necessarily be limited as well. The total collateral requirements for 2024 are shown below in Figure 32.

The average daily collateral total in 2024 was approximately $6.85 billion, marking an 11% decrease from 2023 and breaking the upward trend in ERCOT market collateral levels that had persisted since Winter Storm Uri in 2021. Much of this decline can be attributed to NPRR 1112, which disallows unsecured credit limits and guarantee amounts to be counted as collateral and requires market participants to fully collateralize their obligations with secured instruments such as letters of credit, cash deposits, or surety bonds. This data points to two interesting trends in how market participants manage their collateral requirements.

First, the total amount of collateral posted in ERCOT is consistently higher in aggregate than is strictly necessary for participation in the day-ahead market. Thus, some of the decrease in the recognized forms of collateral did not need to be replaced for market participants to maintain the same degree of participation in the day-ahead market. However, some market participants did have to find alternative sources of collateral, likely increasing their cost to participate in the day-ahead market. There is a cost to posting collateral, whether it involves lost opportunity on cash deposited or the cost of securing and maintaining a letter of credit or surety bond. This increase in cost does not appear to have resulted in an overall reduction in participation in day-ahead market.

Figure 32: Average Daily Collateral Held by ERCOT



C. Point-to-Point Obligations

We cover PTPs in greater detail in Chapter IV, but we discuss them here briefly as they represent a significant share of day-ahead market activity and directly reflect participants' expectations of congestion. PTP obligations are a key part of day-ahead market activity, allowing participants to hedge or speculate on congestion between two locations across the day-ahead and real-time markets. A PTP represents a scheduled flow from a source node to a sink node, with the purchase cost equal to the day-ahead price difference between those nodes times the flow volume. Its value is realized in real time, when the position is liquidated, based on the real-time price difference between the same nodes. The net value of the PTP reflects the change in congestion between the day-ahead and real-time markets. When the payment made to purchase the PTP is lower than the revenue earned in real time, the PTP is profitable for the participant. Figure 33 compares the total day-ahead payments made to acquire these products with the total amount of revenue received by PTP holders in the real-time market over the last three years.

Figure 33: Point-to-Point Obligation Charges and Revenues, 2020-2024

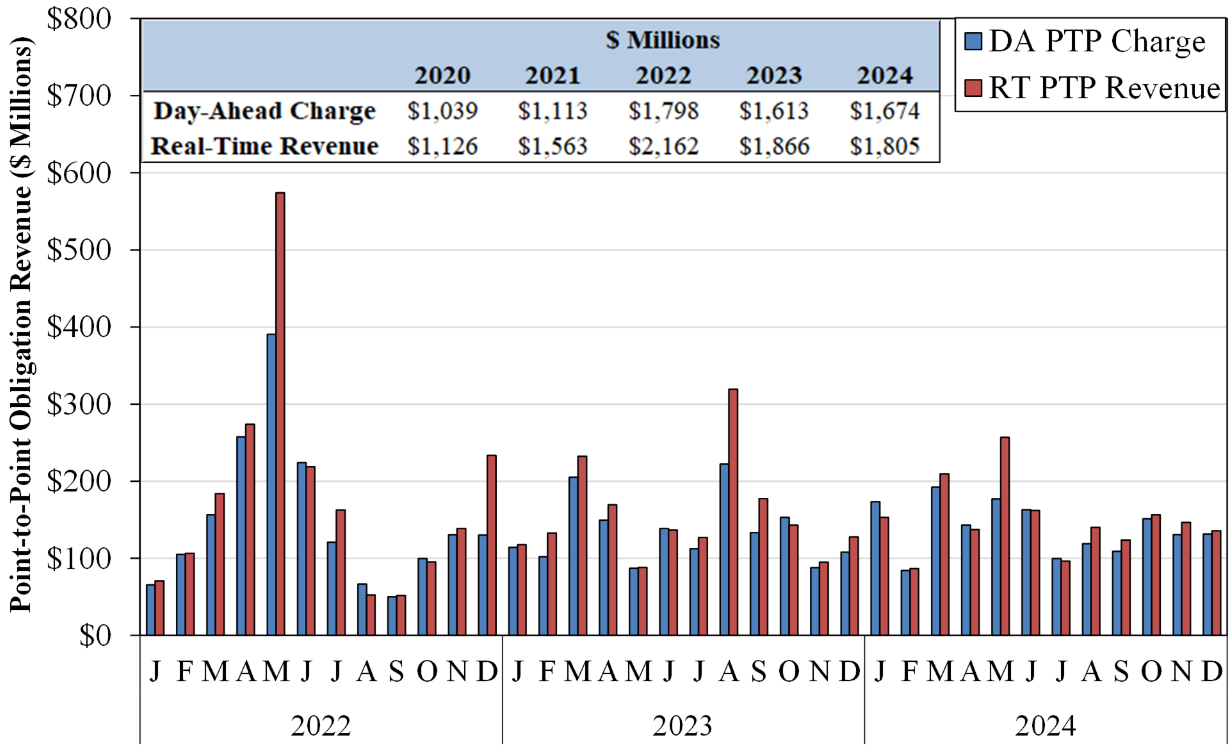
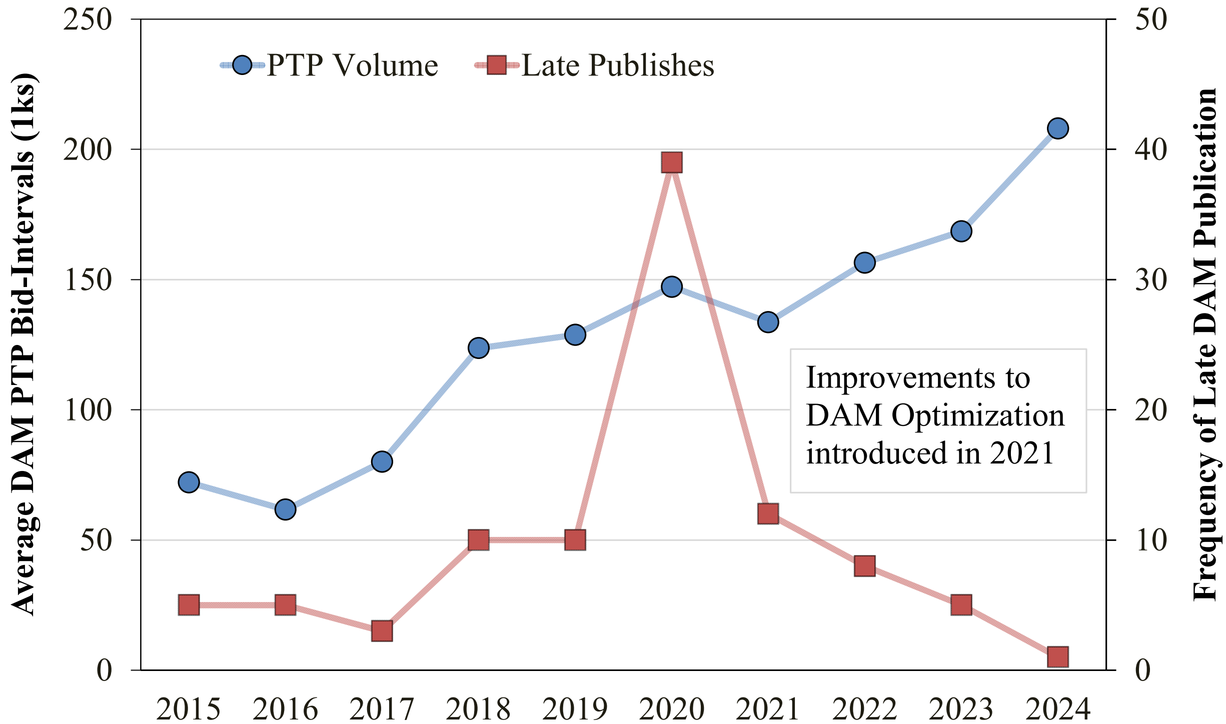


Figure 33 shows that PTPs remained profitable in 2024, paying 7.8% more in real-time than their day-ahead purchase cost. However, the profit margin was about half of what it was in 2023, and the total real-time value of PTPs fell for the second straight year. A key reason for declining profitability is the sharp increase in competition to procure PTPs, as observed via PTP bid volumes that are shown in Figure 34.

The volume[31] of PTP bids has grown by nearly 200% over the past decade, which has implications for both the profitability of PTPs and the performance of the day-ahead market. Higher bid volumes lead to clearing prices that more closely reflect real-time congestion, which reduces the net value of awarded PTPs and limits opportunities for profit. At the same time, the growing number of PTP bids has increased the complexity of solving the day-ahead market optimization, making it more difficult for ERCOT to produce and publish market results within the timeline required by protocol. Many of these PTP bids are unlikely to clear, as their prices do not reflect a realistic expectation of real-time congestion based on recent conditions. However, all PTP bids must be considered when clearing the day-ahead market.

Figure 34: Volume of Day-Ahead Market PTP Bids and Frequency of Late Publication



ERCOT has implemented several improvements to the day-ahead market optimization process in recent years, leading to shorter solve times and a decline in the frequency of late publications. However, continued growth in the volume of PTP bids keeps performance concerns relevant. To address this, we continue to recommend that ERCOT impose a minimum fee on PTP bids. This would discourage participants from submitting bids with a low likelihood of clearing, reduce the overall volume of PTP bids, and help maintain timely and efficient day-ahead market operations.

31 This volume refers to the total number of hourly PTP bid-intervals, including obligations with links to CRR options

D. Ancillary Services Market

Ancillary services sold in the day-ahead market create a physical obligation to provide reserves in real time. Failure to meet this obligation results in penalties that go beyond simply buying back the position. This structure will change once Real-Time Co-optimization (RTC) is implemented, at which point ancillary service awards from the day-ahead market will be financial positions rather than physical obligations. Until then, the ability of the day-ahead market to procure adequate reserves and clear them at competitive prices remains essential. This section summarizes the types of ancillary services purchased in the day-ahead market, the required procurement volumes for each product, and the clearing prices produced by the market.

Ancillary services in the ERCOT market since June 2023 include the following:

- **Regulation Up/Down Service (Reg-Up, Reg-Down).** Regulation reserves include capacity that responds to Load Frequency Control (LFC), which sends out instructions every four seconds to either increase or decrease generation or demand as necessary to keep generation and load in balance from moment to moment and maintain system frequency.
- **Responsive Reserve Service (RRS).** (10-min reserves). Responsive reserves are needed to restore system frequency in the event of rapidly developing contingencies such as unplanned generator outages, rather than for meeting normal load fluctuations. ERCOT procures three different types of responsive reserves: (1) Primary Frequency Response, which all generators have to be able to provide and responds automatically to deviations in system frequency, (2) Under Frequency Relay, which is deployed by tripping NCLRs given a sufficient drop in frequency, and (3) Fast Frequency Response, which is provided by energy storage resources (ESRs) that can respond to deviations in frequency within 30 cycles or by NCLRs.
- **ERCOT Contingency Reserve Service (ECRS).** (10-min reserves) The latest addition to ERCOT's suite of ancillary services, the purpose of ECRS is to restore frequency within 10 minutes of a significant frequency deviation and for recovery of deployed regulation service, to compensate for intra-hour net load forecast uncertainty and variability on days in which large amounts of online thermal ramping capability is not available, or to compensate for times during which there is a limited amount of capacity available to SCED. ECRS can be provided by online and offline units.
- **Non-Spin Reserve Service (NSRS).** (30-min reserves) Similar to ECRS, NSRS are needed to compensate for intra-hour net load forecast uncertainty that results in under commitments of capacity or inefficient dispatch instructions and can be provided from online and offline units.

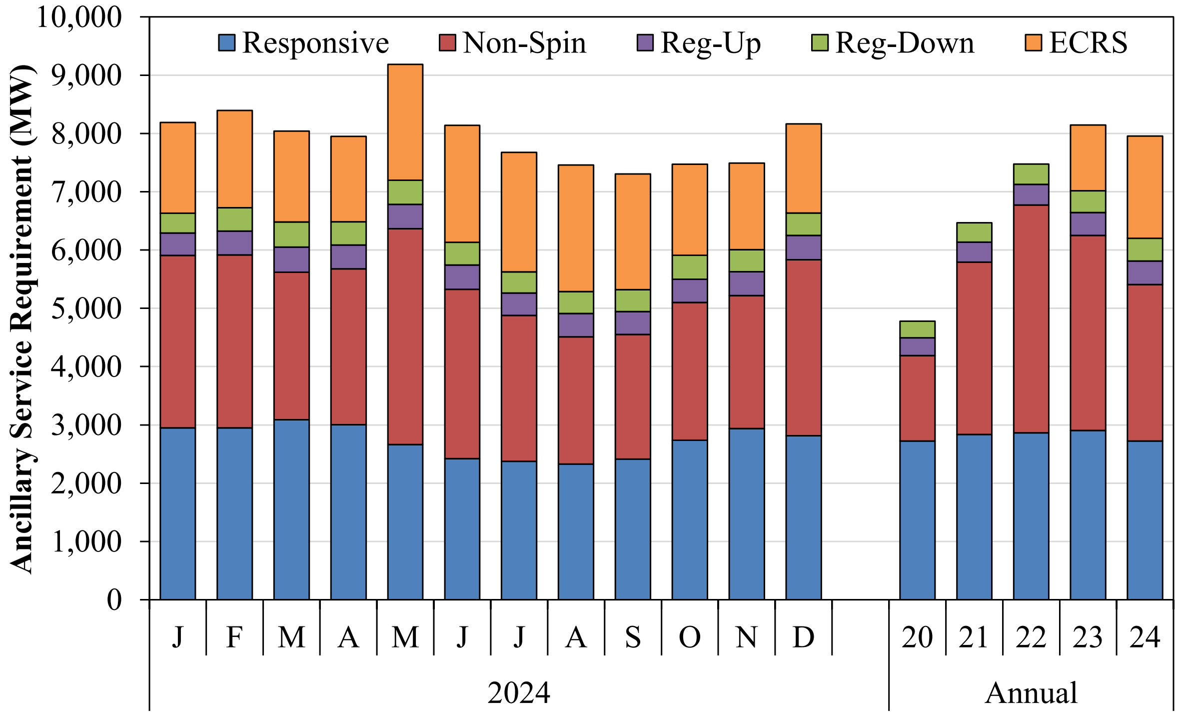
1. Ancillary Services Requirements

The volume of each ancillary service that ERCOT procures is determined on a month-hour basis according to the corresponding AS Methodology.[32] The resulting schedule, called the AS Plan, is shown for 2024 on a monthly basis and for 2020-2024 on an annual basis in Figure 35. This chart shows that the average annual volume of AS procured has increased substantially over the last five years, though it decreased by approximately 2.3% from 2023 to 2024, mainly due to the following changes to the AS Methodology:

- The floor on RRS volumes for peak hours was reduced from 2,800 MW in 2023 to 2,300 MW in 2024.
- For January through May of 2023, the NSRS methodology used the ten hour-ahead forecast error, but the six hour-ahead forecast error was used for the remainder of 2023 and all of 2024.

The month with the largest average volume in the AS Plan in 2024 was May, which is driven by volatility in the net load forecast.

Figure 35: Average Ancillary Service Capacity by Month



Despite the reduction in the AS Plan from 2023 to 2024, ERCOT still procures a greater volume of operating reserves compared to other Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs), as shown in Figure 9 in Chapter I. Under the current

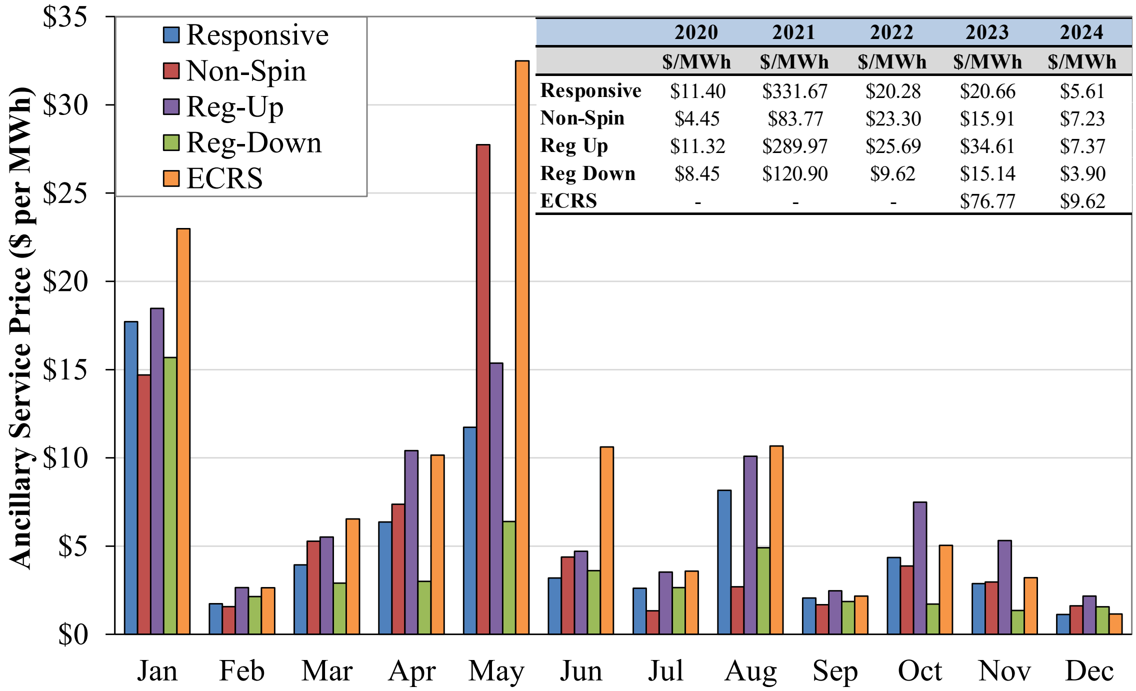
32 https://www.ercot.com/files/docs/2022/06/07/ERCOT-Methodologies-for-Determining-Minimum-AS-Requirements-040125.zip

market design, the DAM will clear the entire AS plan, often resulting in high clearing prices above the marginal reliability value associated with each product. This excess cost is paid by consumers; specifically, ERCOT allocates the cost of procurement to Load Serving Entities (LSEs) on the basis of their real-time adjusted metered load. Market participants representing qualified resources may self-schedule ancillary services using those resources to reduce their exposure to the cost of ancillary services procured in the day-ahead market.

2. Ancillary Services Prices

Figure 36 presents the monthly average clearing prices of capacity for the five ancillary services in 2024, and the inset table shows the average annual prices over the last five years. This highlights a notable spike in ancillary service prices in May 2024, which was driven primarily by extreme market conditions on May 8th and May 26th. On both days, unusually hot weather created heightened uncertainty around wind and solar output, which in turn led to sharp increases in day-ahead prices. While January also saw elevated prices, the overall trend for 2024 shows a substantial decline in average ancillary service prices compared to 2023, most notably for ECRS, which dropped from $76.77 per MWh to $9.62 per MWh. The total cost of ancillary services per MWh of load dropped by approximately 74%, as shown in Figure 1 in Chapter I.

Figure 36: Ancillary Service Prices, 2020-2024



	2020	2021	2022	2023	2024
	$/MWh	$/MWh	$/MWh	$/MWh	$/MWh
Responsive	$11.40	$331.67	$20.28	$20.66	$5.61
Non-Spin	$4.45	$83.77	$23.30	$15.91	$7.23
Reg Up	$11.32	$289.97	$25.69	$34.61	$7.37
Reg Down	$8.45	$120.90	$9.62	$15.14	$3.90
ECRS	-	-	-	$76.77	$9.62

Like with the decrease in energy prices, the decrease in ancillary service prices can largely be explained by two factors. There were fewer days of extreme temperatures in summer 2024 compared to 2023, and there was considerably more solar and energy storage capacity online. Both factors contributed to an increased supply of reserves that resulted in lower ancillary service prices on average.

IV. Transmission Congestion and Congestion Revenue Rights

An essential function of any electricity market is to efficiently manage power flows on the transmission networks. Congestion management occurs as the markets coordinate the dispatch of generation to ensure that the resulting power flows do not exceed the operating limits of the transmission facilities. This coordination occurs through the real-time market dispatch model, i.e., SCED, which schedules generation to meet demand based on each generator's energy offer curve and their corresponding impact on transmission constraints. The result of this market dispatch is a set of locational prices that vary across the network and resulting congestion costs that are collected from participants.

Persistent transmission congestion plays a central role in shaping electricity prices across the grid. In 2024, real-time operations experienced at least one binding constraint 86% of the time, meaning power flows on those constraints met or exceeded their rated operating limits. This widespread congestion leads to two fundamental pricing outcomes, the most prominent being locational marginal prices (LMPs) which represent the variation in electricity prices across different geographic locations. LMPs reflect the cost to serve load at each node while respecting the physical limits of the transmission network. These price differences are not arbitrary; they are designed to guide economic decisions. Higher LMPs in constrained areas signal generators to site where they can relieve congestion and earn higher revenues, while discouraging load from locating in areas that exacerbate transmission constraints.

Another outcome of congestion is a spread between the prices paid by load and received by generation. The logic of locational marginal pricing tends to result in higher prices for load nodes than for generation nodes. The difference between the total bills paid by load and the total revenue earned by generators is called congestion rent. A fundamental aspect of electricity market design is the allocation of this congestion rent. This refers to determining who receives the excess revenue collected from load after generators have been paid, while ensuring the system operator remains revenue neutral. In ERCOT, the allocation of congestion rent collected from the day-ahead market (DAM) is made according to congestion revenue rights (CRRs).

CRRs are economic property rights funded by congestion rent collected in the day-ahead market. The owner of a CRR is entitled to a share of this congestion rent based on the price difference between two locations: the source, where power is injected into the system, and the sink, where power is withdrawn. CRRs are purchased through an auction and are defined by paths, each consisting of a specific source and sink. They are denominated in megawatts, and the volume of CRRs available on each path is limited by the physical transmission capacity between the two locations. CRRs can be purchased in monthly blocks up to three years in advance.

This chapter summarizes congestion costs and revenues in 2024. We first discuss the value of congestion in the day-ahead and real-time markets, which totaled approximately $1.9 billion each. We then discuss the dynamics of the CRR markets in 2024, followed by a summary of real-time congestion shortfalls. Finally, we discuss how ERCOT's load zones could be updated to better reflect congestion patterns that have changes significantly since they were introduced.

A. Day-Ahead Congestion

The day-ahead market produces financially binding schedules for supply, demand, and point-to-point (PTP) transactions that account for transmission system limits. When these limits are binding, congestion leads to a more expensive dispatch solution and causes energy prices to vary across locations in the network. Congestion in the day-ahead market is influenced by factors such as planned transmission outages and forecasts for load and intermittent renewable generation. These factors shape how market participants hedge their positions before real time.

Figure 37 illustrates how congestion is valued in the day-ahead market and highlights the role of Generic Transmission Constraints (GTCs[33]) in managing system stability. Day-ahead congestion values are calculated as the product of power flows over each constraint and the constraint's shadow price, which reflects the marginal economic cost of that constraint.

Figure 37: Value of Day-Ahead Congestion by Zone

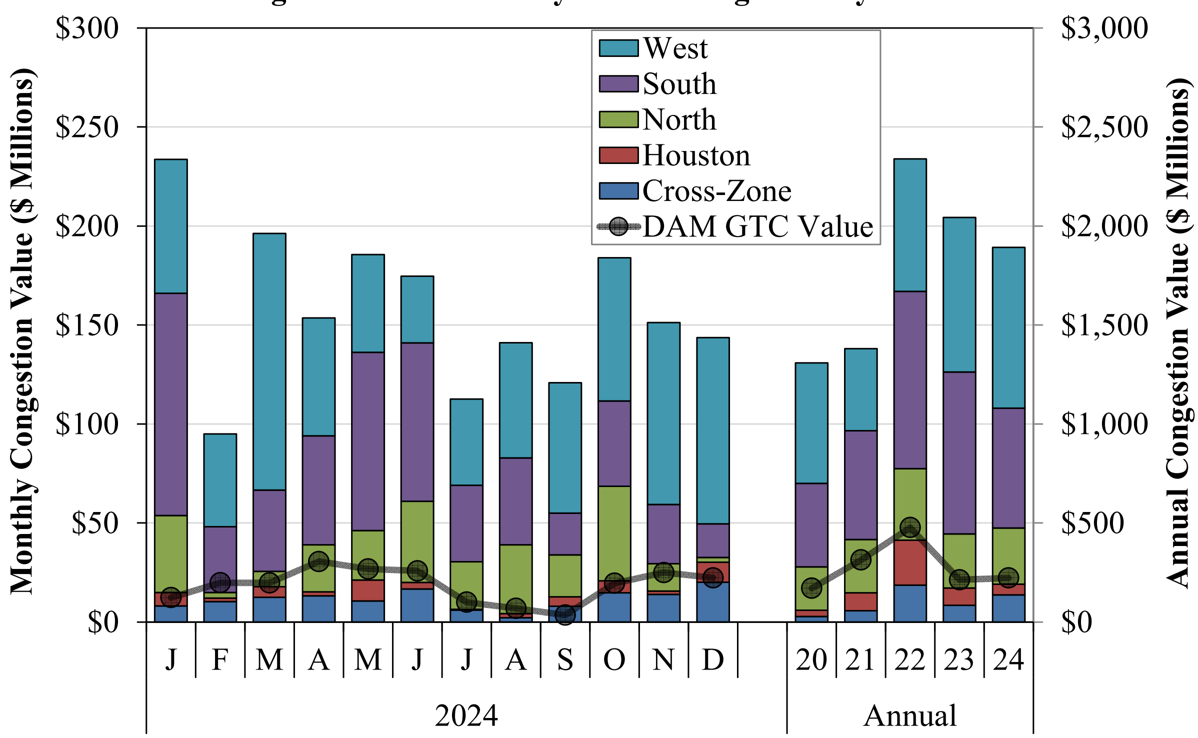


Figure 37 presents congestion values both within zones and across zones, and includes congestion linked to GTCs. These constraints are used to manage grid stability across

33 A GTC is a transmission constraint made up of one or more grouped Transmission Elements that is used to constrain flow between geographic areas of ERCOT for the purpose of managing stability, voltage, and other constraints that cannot otherwise be modeled directly in ERCOT's power flow and contingency analyses and are based on offline studies (i.e., Real-Time Contingency Analysis (RTCA) will not indicate concerns).

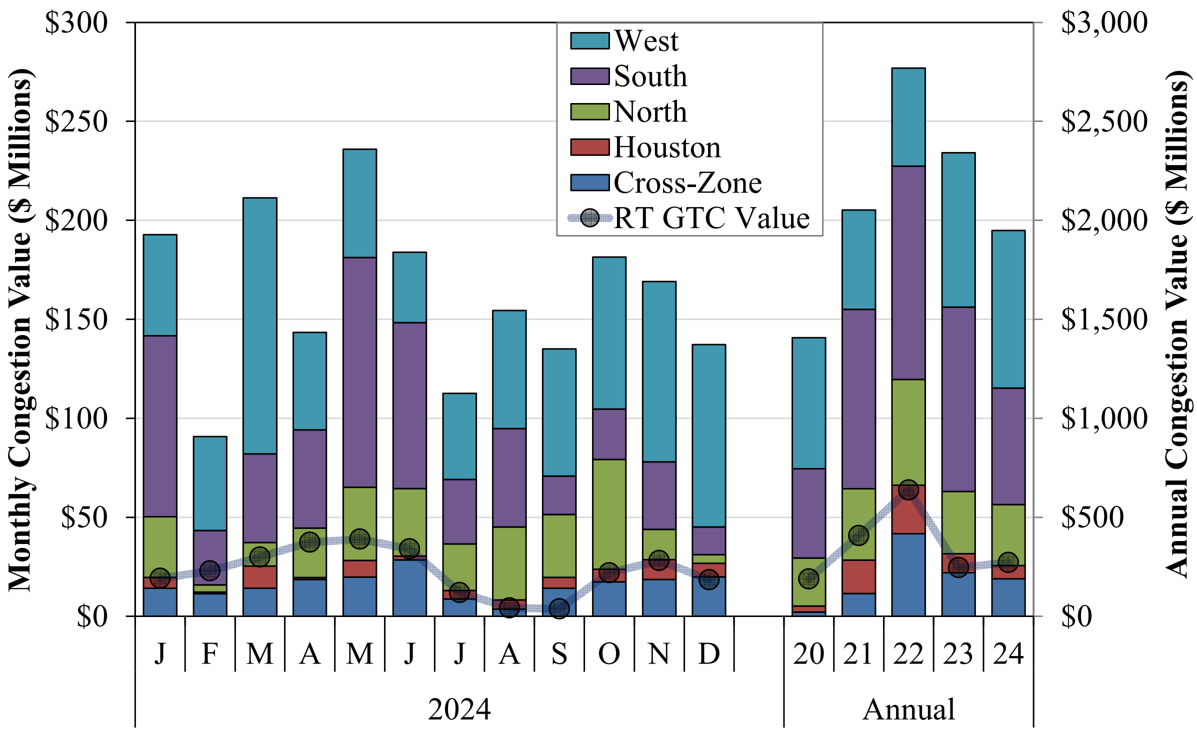
geographic regions, rather than enforce thermal transmission limits. GTCs are especially relevant in areas with significant renewable generation, which are often far from load centers and can contribute to system stability concerns.

The data in Figure 37 also shows that day-ahead congestion declined throughout 2024, continuing a downward trend that began after a peak in 2022. This suggests improving transmission conditions or more effective congestion management in the market. The figure also indicates that roughly 12% of the total congestion value in the day-ahead market during 2024 was attributable to GTCs.

B. Real-Time Congestion

While the expected costs of congestion are reflected in the day-ahead market, physical congestion occurs only in the real-time market, and ERCOT operators manage power flows across the network as physical constraints become binding in real time. Unexpected changes in system conditions between the day-ahead and real-time markets often lead to differences in congestion values. These changes can include net load forecast errors, forced outages, and other deviations from expected conditions. Figure 38 summarizes the monthly real-time congestion for 2024 as well as annual values from 2020 through 2024.

Figure 38: Value of Real-Time Congestion by Zone



Real-time congestion in 2024 followed similar trends to the day-ahead market. Overall congestion was approximately 3% higher in the real-time market than in the day-ahead market, corresponding with higher real-time congestion across zones and within the Houston and North zones. There was also approximately 22% more congestion associated with GTCs in the real-time market, likely a result of forecast error for generation from renewables. Several new transmission components were energized in the South Zone in 2024 that contributed to the decrease in intra-zonal congestion compared to 2023.

1. Types and Frequency of Constraints in 2024

There are thousands of transmission constraints associated with physical elements in the ERCOT transmission network. To make the dispatch problem solved by SCED more computationally tractable, ERCOT only activates a subset of these constraints at any given time. In addition to GTCs, the limits of which are set prior to the operating day, ERCOT also activates constraints based on the Real-Time Contingency Analysis (RTCA) process that runs on an ongoing basis. The RTCA evaluates network flows under many contingency scenarios and is used to determine when a constraint is at risk of being violated and needs to be activated in SCED. Constraints are considered binding when dispatch costs are incurred to maintain transmission flows below the constraint limit, and not binding when they do not require a redispatch of generation and have no effect on prices. Figure 39 summarizes the frequency of active and binding constraints during 2024. The bars show the percentage of time (y-axis) at different load levels and annually (x-axis) with a binding or active constraint. The green line denotes the average number of constraints at different load levels.

Figure 39: Frequency of Binding and Active Constraints by System Load Level

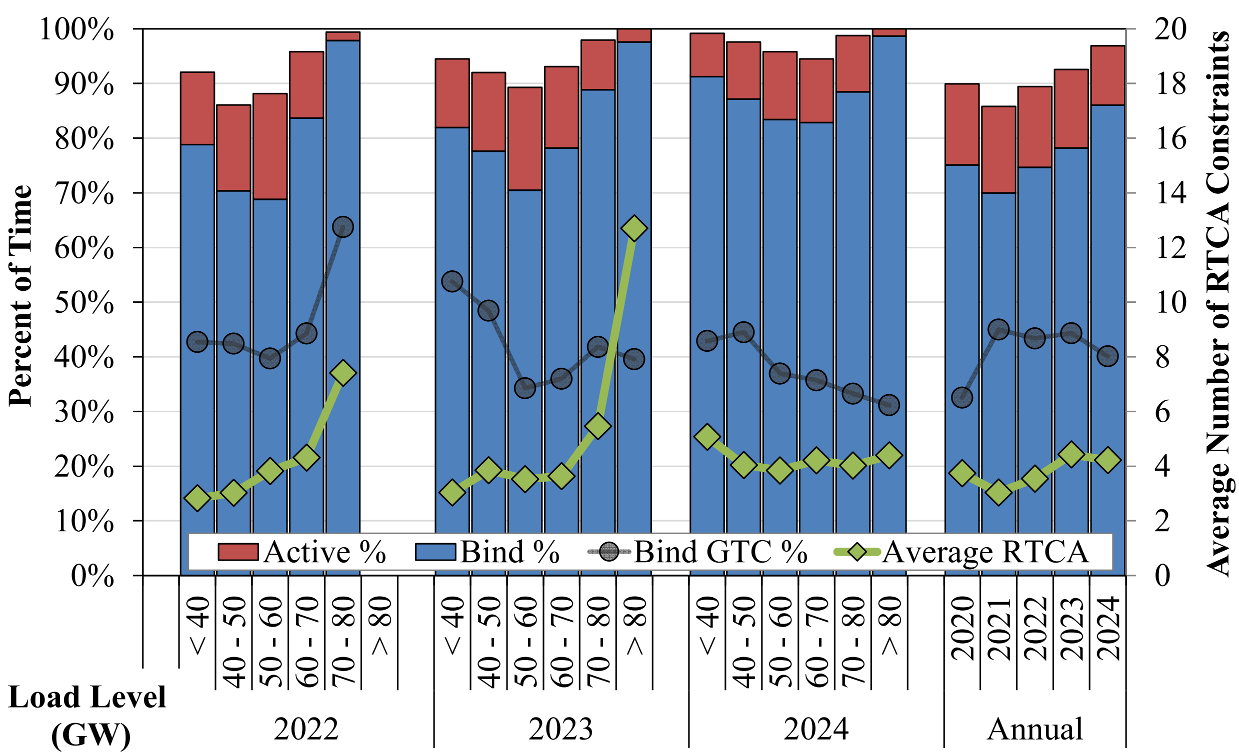


Figure 39 shows the following:

- The ERCOT system had at least one binding constraint 86% of the time in 2024, an increase from 78% in 2023 and 75% in 2022.
- Consistent with previous years, the average number of active constraints was lowest when load was in the range of approximately 50 GW.
- Similar to 2023, the percentage of the time in which a GTC was binding decreased as the load level increased.
- Unlike in 2022 and 2023, there was not a noteworthy increase in the number of constraints flagged by RTCA when load was above 80 GW.

2. Violated Constraints

The shadow price of a constraint represents the marginal cost of redirecting the flow of energy around a binding constraint. A constraint is considered violated when the market dispatch flows exceed the transmission limit for the constraint. Such violations impose reliability costs or risks on the system that are embedded in the shadow price caps used by ERCOT to dispatch the system and set prices.[34] When the marginal costs of procuring relief through the market dispatch exceeds the reliability costs of violating the constraint, the shadow price caps will: a) prevent the market from incurring additional dispatch costs; and b) set the shadow price for the constraint, which determines the congestion prices at locations that affect the violated constraint.

The shadow price caps during 2024 were:

- $5,251 per MW for base-case (non-contingency) constraints or voltage violations,[35]
- $4,500 per MW for 345 kV constraints,
- $3,500 per MW for 138 kV constraints, and
- $2,800 per MW for 69 kV thermal violations.
- GTCs are considered base-case stability constraints (for voltage or transient conditions) with a shadow price cap of $5,251 per MW.

Note that ERCOT implemented a process through NPRR 1230 to increase the shadow price cap on base-case constraints in response to IMM recommendation 2023-1. Increasing the shadow price cap allows SCED to produce more expensive dispatch solutions to avoid violating these

34 See Methodology for Setting Maximum Shadow Prices for Network and Power Balance Constraints.

35 OBDRR 037, *Power Balance Penalty and Shadow Price Cap Updates to Align with PUCT Approved High System-Wide Offer Cap*, reduced the shadow price cap for base-case constraints from $9,251 per MW to $5,251 per MW effective April 1, 2022.

constraints. Figure 40 shows the distribution of violated constraints at the various violated constraint overload percentages since 2020.

Figure 40: Overload Distribution of Violated Constraints

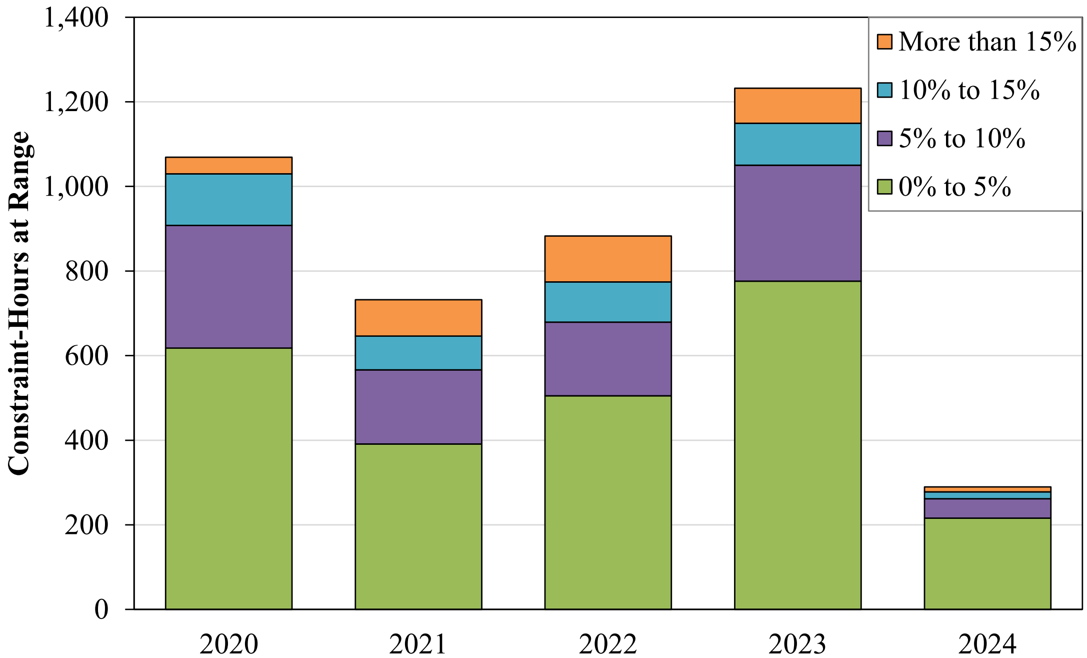
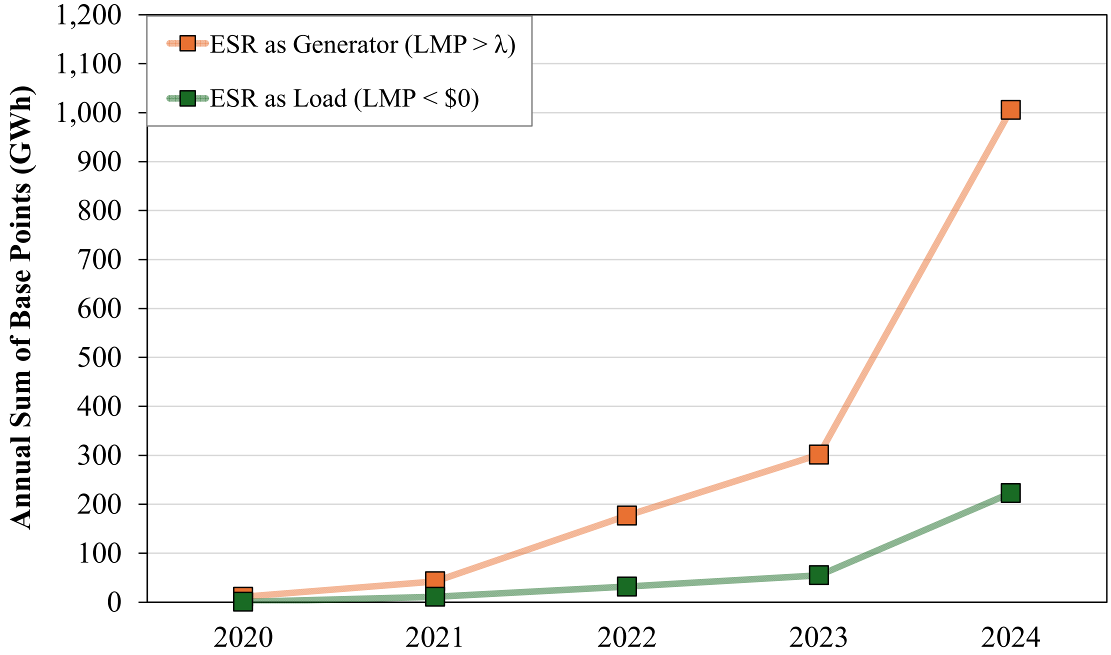


Figure 40 shows that the overall rate of violated transmission constraints was down considerably in 2024 from prior years with only 290 constraint-hours of violations compared to an average of 979 constraint-hours of violations for 2020-2023. In 2024, ERCOT recorded one of the lowest numbers of constraint-hours of violations since the nodal market began in 2011. The only year with fewer was 2016, which had 271 constraint-hours of violations. The sharp decline may be partly explained by the growing participation of energy storage resources (ESRs) in the real-time energy market. As the ancillary services market has become saturated, a larger share of ESR revenue depends on energy arbitrage, accelerating this shift toward real-time market activity.

ESRs are particularly well-suited to help resolve congestion, because they can act as both supply and demand. When congestion can be resolved by increasing load at a particular location, ESRs can often charge at prices below $0. Conversely, when congestion can be resolved by increasing supply at a particular location, ESRs can discharge and earn higher prices to reflect the value of serving load subject to binding transmission constraints. To illustrate this trend, we show the annual total of base points awarded to ESRs helping congestion as generators or as loads in Figure 41. For ESRs that act as generators, we aggregate the sum of base points awarded with an LMP greater than the system lambda, indicating that injections of energy at that location tend to help manage congestion. For ESRs acting as loads, we aggregate the sum of base points awarded with an LMP less than zero, meaning the benefit of ESR load in resolving congestion is valuable enough to the system that they are paid to charge. Figure 41 shows an accelerating increase in both trends over the last five years.

Figure 41: Annual Sum of Base Points Awarded to ESRs Helping Congestion



As in previous years, a large majority of constraint violations in 2024 were less than or equal to 5% above the limit of the constraint. These relatively small violations are priced at the same shadow price cap as the more severe violations. This raises some concerns because the use of a single shadow price cap causes the pricing of the violations to not vary with the severity of the violation. Hence, it may be advisable to reconsider implementing transmission demand curves, which would recognize that the reliability risk of a post-contingency overload increases as the violation amount increases. Small violations should have lower shadow prices than large violations. The IMM filed a revision request to implement transmission constraint demand curves, which was ultimately withdrawn in 2022 for lack of support.

C. CRR Market Outcomes and Revenue Sufficiency

As discussed in the introduction to this chapter, CRRs are economic property rights entitling the holder to the day-ahead congestion payments or charges between two locations. In this section, we discuss the timeline and inputs to the CRR auctions, the allocation of the revenues from the CRR auctions to load, profitability trends for CRRs, and the funding of CRRs through day-ahead market congestion rent.

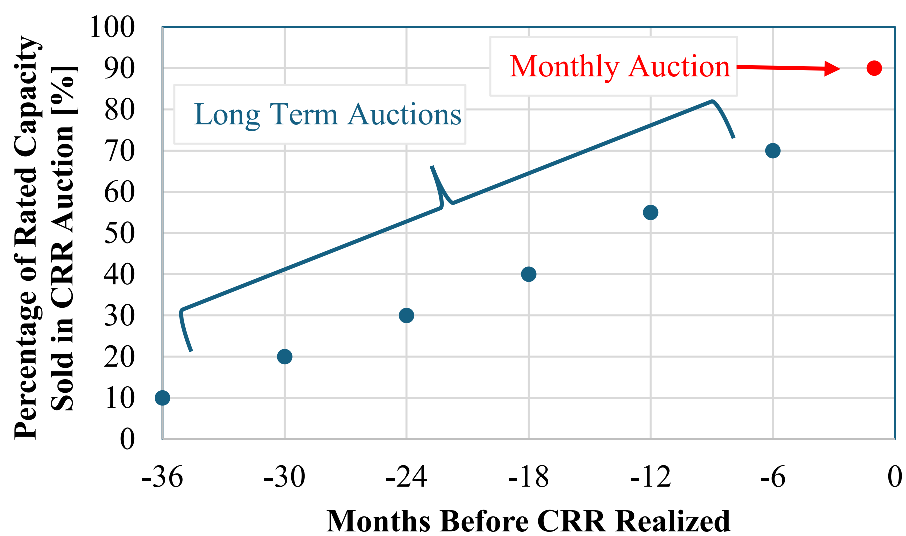
1. CRR Auction Timeline

A CRR is a financial product that reflects the value of transmitting energy from one location to another. In the CRR auction, each path is modeled as an injection of energy at the source and a withdrawal at the sink. CRRs can be purchased either as obligations or as options. The owner of a CRR obligation is entitled to the difference in price between sink and the source produced in the day-ahead market, but they are also obligated to pay the difference of these prices if the price at the source is greater than the price at the sink. The owner of CRR options, on the other hand, is entitled to the positive price differences between the sink and the source but is not obligated to pay for a negative price difference. A subset of CRRs called Pre-Assigned Congestion Revenue Rights (PCRRs) are allocated to Non-Opt-In Entities (NOIEs) based on generation units owned or contracted prior to the start of retail competition. Parties receiving PCRRs pay only a fraction of the auction value of a CRR between the same locations.

Each CRR remains valid for the full duration of the month or block of months for which it is purchased. CRRs are sold through two recurring types of auctions: monthly auctions and long-term auctions, the latter held twice each year. In the long-term auction, CRRs can be bought either for individual months covered in that auction or as a block of consecutive months within that auction. The long-term auctions are run as far as three years (36 months) in advance, and then subsequent long-term auctions are run every six months. The last long-term auction runs six months in advance of real-time. The monthly auctions are run the month before the CRRs are realized, i.e., the monthly auction for March runs in February.

Running CRR auctions years in advance of the real-time market allows for more hedging opportunities for market participants and informs forward price formation. That said, there is also more uncertainty in forecasting future congestion, which is a function of load, generation and transmission outages, and generation from renewable resources. This uncertainty creates the potential that payments to CRR account holders (CRRAHs) could exceed congestion rent collected in DAM and, therefore, increases the risk of a shortfall in payments to CRRAHs. To avoid such shortfalls, ERCOT only sells a percentage of the rated capacity for each CRR path in the long-term auctions and increases that percentage in each subsequent long-term auction through the final monthly auction as shown in Figure 42.

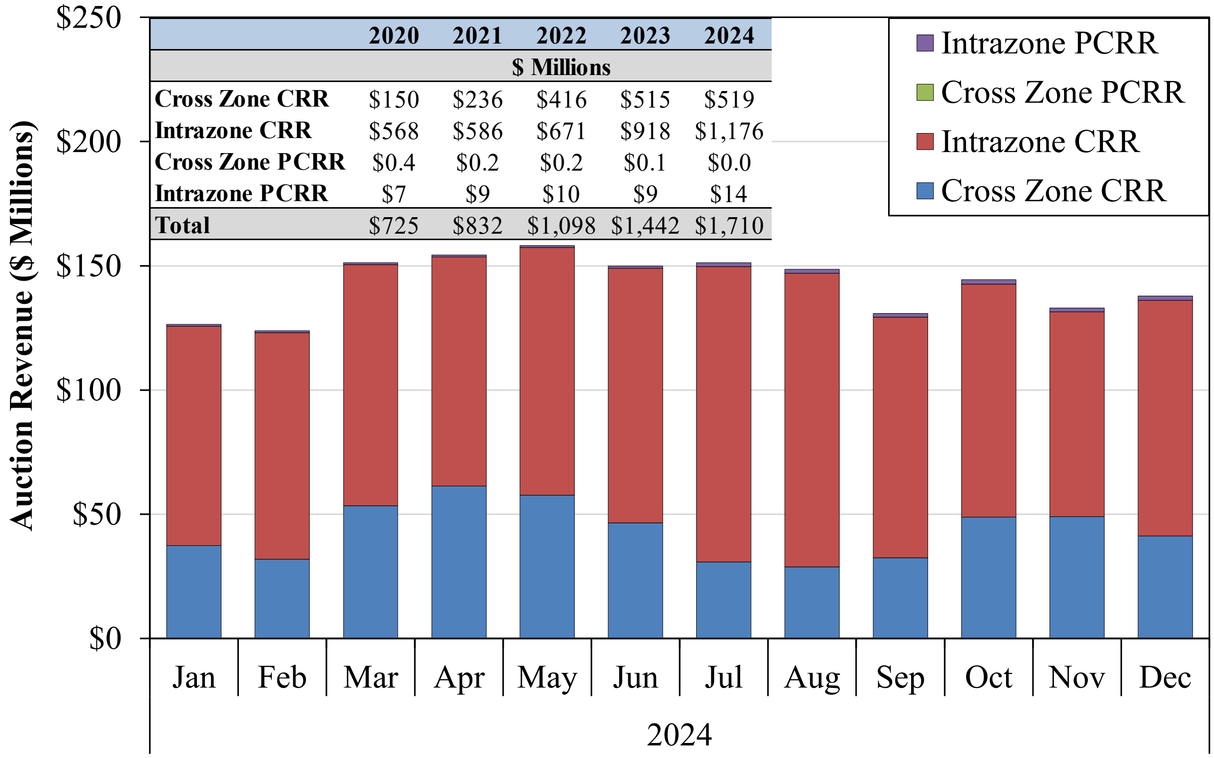
Figure 42: Schedule for CRR Capacity Sold in Long-Term Auctions



2. CRR Auction Revenues

The total amount of CRR auction revenue increased by almost 19% from 2023 to 2024, continuing a consistent trend of increasing annual CRR auction revenue despite decreasing congestion since 2022, as shown in Figure 43.

Figure 43: 2024 CRR Auction Revenue, 2020 - 2024



	2020	2021	2022	2023	2024
	$ Millions				
Cross Zone CRR	$150	$236	$416	$515	$519
Intrazone CRR	$568	$586	$671	$918	$1,176
Cross Zone PCRR	$0.4	$0.2	$0.2	$0.1	$0.0
Intrazone PCRR	$7	$9	$10	$9	$14
Total	$725	$832	$1,098	$1,442	$1,710

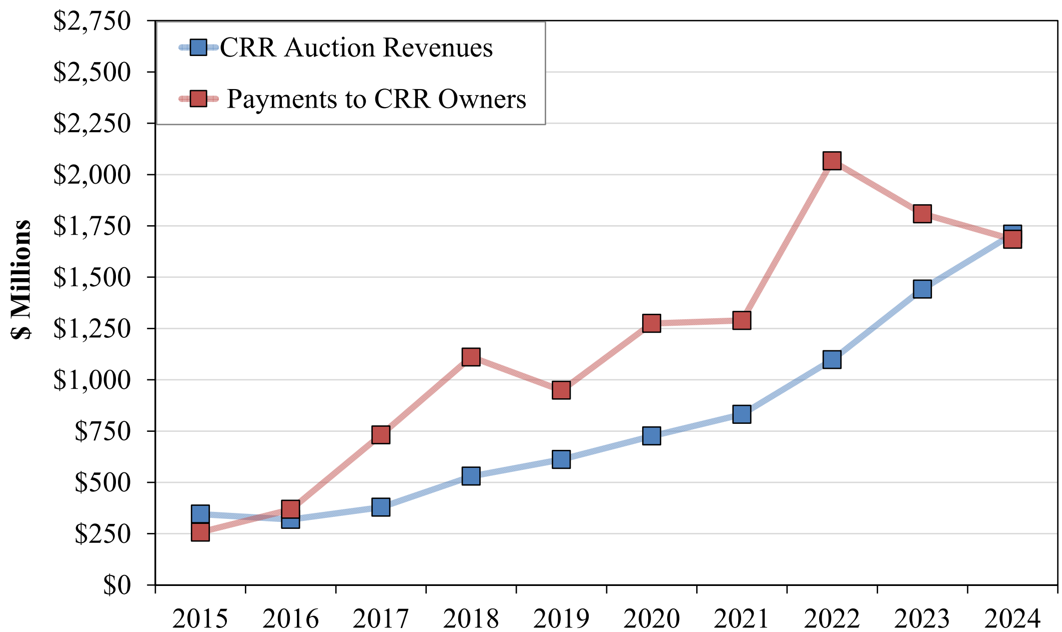
The revenues ERCOT receives by selling CRRs are distributed to Load Serving Entities (LSEs) according to the CRR Auction Revenue Distribution (CARD) process. Revenues from cross-zone CRRs are allocated to LSEs ERCOT-wide based on their system-wide load ratio share during the coincident peak interval for each month. Revenues from CRRs that have the source and sink in the same zone are allocated to loads within that zone based on their zonal load ratio share during the same coincident peak interval.

As the revenues from CRR purchases and corresponding distributions of CARD payments have increased, this methodology has garnered increased scrutiny, as it potentially creates adverse economic incentives for loads to increase their consumption during periods of high demand, when high energy prices should incentivize loads to decrease their consumption. To date, the only clear examples of such behavior have come from DC tie operators. To remove these adverse incentives for DC tie operators, ERCOT implemented NPRR 1030 to allocate CARD payments for DC ties based on their monthly load ratio shares. This methodology severely reduces the incentive for DC ties to increase their consumption for the sake of increasing their allocation of CARD revenue.

3. CRR Profitability

Figure 44 shows annual aggregate CRR auction revenue and payments to CRR owners, the difference of which represents CRR profitability. Overall, CRRs have been profitable on an annual basis since 2016. In 2024, however, CRRs paid out a loss of approximately 1.5%. This decrease in profitability was a continuation of a trend started in 2023 where the margins on CRR profits decreased to 25% from 88% in 2022. This reduction in what had been exceptionally high profit margins suggests that CRR pricing has become more efficient in recent years. That is, the bids in CRR auctions are increasingly reflecting the ultimate value of CRRs.

Figure 44: CRR Auction Revenue, Payments, and Congestion Rent



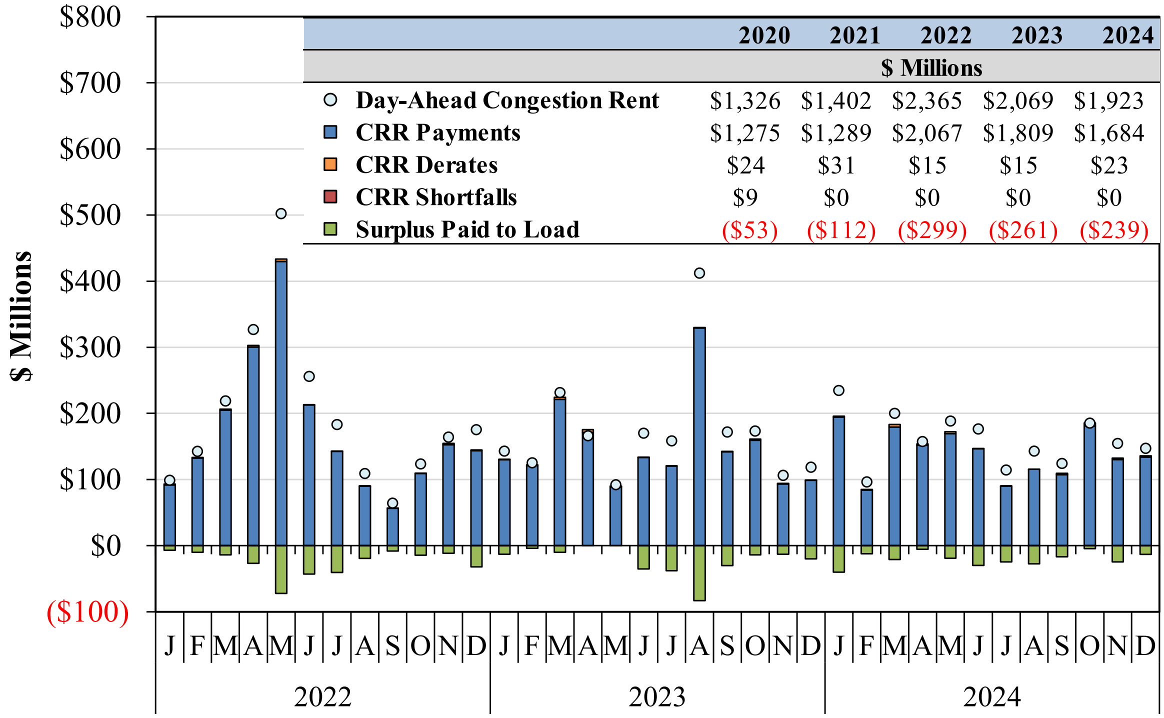
4. CRR Funding Levels

The target value of a CRR is the product of its quantity and the price difference between sink and source, which reflects the entitlement to the holder. It is important for the integrity of the CRR market that the CRRs are fully funded by the congestion rent. ERCOT will only pay less than the target value when the day-ahead congestion rent is insufficient, which can occur when the network flows modeled in the CRR auction are greater than the flows in the day-ahead market. This is generally the result of unforeseen outages or other factors that reduce the transmission capability between the CRR auction and the DAM.

Settlement of CRR Shortfalls. When a shortfall occurs on a specific transmission facility due to oversold flows, payments to CRRs that sink at generator locations affecting that facility will be reduced. These reductions are based on the decrease in day-ahead transfer capability. If revenue is still insufficient after this adjustment, the remaining shortfall is shared across all holders of positively valued CRRs through a prorated charge, reducing their overall payments.

Settlement of CRR Surpluses. When day-ahead congestion rent exceeds the amount owed to CRR holders, the excess is tracked in a monthly settlement process called the balancing account. ERCOT uses this excess congestion rent to repay CRR holders who were previously assessed shortfall charges, effectively refunding those amounts. If there is not enough excess congestion rent in the current month, the rolling CRR balancing fund from prior months can be used to fully pay CRR holders. The CRR balancing fund has a $10 million cap, beyond which ERCOT disperses the remaining amount to LSEs. CRRs were fully funded in every month of 2024, and no short payments occurred. Figure 45 shows monthly CRR surpluses and shortfalls since 2020.

Figure 45: CRR Solvency and Surplus Payments to Load



In 2024, the total day-ahead surplus was approximately $239 million, which was an 8.9% decrease from 2023. Despite this decline, congestion rent from the day-ahead market was sufficient to fully fund CRRs in every month of the year. As a result, the balancing account remained capped at $10 million, and all surplus congestion rent above that threshold was returned to LSEs. The last CRR short payment occurred in November 2020. ERCOT's practice of offering only 90% of the forecasted transmission capability in CRR auctions reduces the likelihood of future funding shortfalls.

Importantly, even though the day-ahead market produced sufficient revenues to fully fund the CRRs, many CRRs were derated in 2024 because of the mandatory deration process. In total, CRR deratings resulted in a $23 million reduction in payments to CRR holders. These deratings reduced ERCOT's overall funding percentage to approximately 99%, comparable to the previous year. Derating CRRs when the market is producing sufficient revenue introduces unnecessary risk to those buying CRRs, which could ultimately result in lower CRR auction revenues.

D. Real-Time Congestion Shortfalls

Just as reductions in network capability from the CRR auctions to the DAM can result in CRR shortfalls, reductions in the network capability between the day-ahead market and the real-time market can result in real-time congestion shortfalls. In addition to outages or limit changes, binding real-time constraints that are not modeled in the day-ahead market can produce real-time congestion shortfalls. Shortfalls are costs incurred by ERCOT to lower the real-time flows when day-ahead scheduled flows exceed the flows the network can support in real time. These real-time congestion shortfall costs are paid for by charges to LSEs as part of the uplift charge known as the Revenue Neutrality Allocation (RENA).

RENA exists to ensure that ERCOT remains revenue-neutral, which means payments equal charges. In general, RENA uplift occurs when there are differences in power flow modeling between the day-ahead and real-time markets, including:

- Transmission network modeling inconsistencies between the day-ahead and real-time market (model differences);
- Differences between the load distribution factor (LDF) used in the DAM and the actual real-time load distribution (LDF contribution);
- Day-ahead PTP obligations linked to options[36] settlements (CRR uplift);
- Extra congestion rent that accrued when real-time transmission constraints were violated (overflow credit); and
- Other factors, including the price floor in the real-time market at -$251 per MWh (Other).

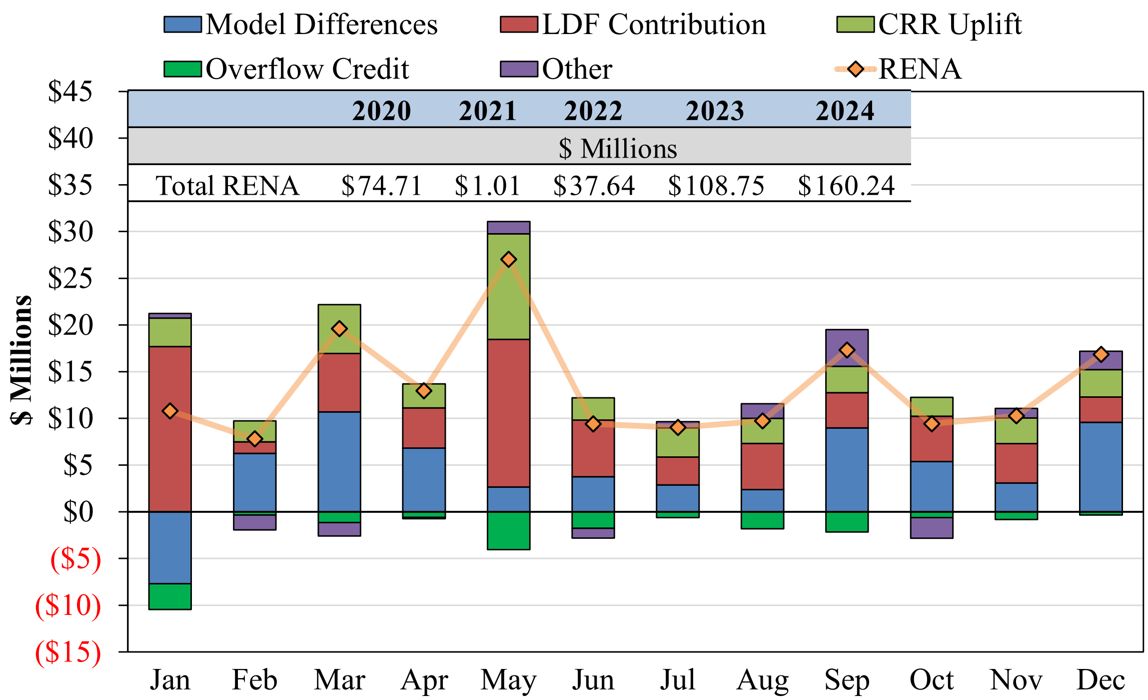
Figure 46 provides an analysis of RENA uplift in 2024, separately showing the components of RENA on a monthly basis. Net negative uplift represents a net payment to load. RENA uplift grew to total $160 million in 2024, up from $109 million in 2023.

Figure 46 shows that the largest positive contributor to RENA uplift in 2024 was the LDF Contribution totaling $75 million. Uplift associated with differences in the transmission models between the day-ahead and real-time markets was also one of the largest factors, accounting for $55 million in RENA uplift. It is a non-trivial task to maintain accurate and consistent LDFs across all markets, particularly in areas with large amounts of localized load growth. To the extent ERCOT is unable to predict accurate LDFs across all markets, RENA impacts will persist. NPRR 1004, *Load Distribution Factor Process Update,* approved on August 11, 2020, is still pending implementation, but should reduce this uplift.[37] This change will introduce load forecast models to calculate daily LDF rather than the current seasonal LDF based on historical patterns.

[36] A PTP obligation linked to an option (PTPLO) is a type of CRR that entitles a NOIE's PTP Obligation in the day-ahead market to reflect the NOIE's PTP Option that it acquired in the CRR auction or allocation. Qualified PTPLOs are modeled as obligations but settled as if they were options.

[37] NPRR 1004, Load Distribution Factor Process Update.

Figure 46: Factors Contributing to RENA, 2020-2024



We encourage ERCOT to seek continuous improvement in aligning the transmission models between the day-ahead and real-time markets. This is a challenge for all wholesale market operators but must be a high priority because it facilitates efficient day-ahead market performance and eliminates opportunities for participants to extract rents associated with differences that ultimately raise the RENA uplift and the costs to ERCOT's consumers.

E. Load Zone Configuration

LMPs are calculated at individual generation and load nodes. Loads are settled according to their corresponding Load Zone price, which is calculated as the load-weighted average of the load node LMPs within each zone. Settling loads according to zonal prices reduces uncertainty and volatility caused by congestion, making it easier for loads to hedge their positions and manage their cash flow. However, for zonal pricing to send efficient signals to load for consumption and investment, it is crucial that the load zones accurately reflect the topology of the network. That is why the IMM recommended in the 2020 SOM that ERCOT update the configuration of load zones and introduce new load zones to recognize key transmission constraints and minimize intra-zonal congestion.

The case for improving the current load zone configuration has grown stronger in recent years. Renewable development, especially solar, has expanded rapidly, while new patterns of load growth have become increasingly misaligned with the existing load zone map. The Permian Basin, a major hub for oil and gas production, has emerged as a high-cost load pocket. In contrast, the Texas Panhandle often sees negative prices due to frequent curtailment of abundant wind generation. Placing both of these regions within the same load zone (West) leads to inefficient pricing that fails to reflect the underlying differences in system conditions.

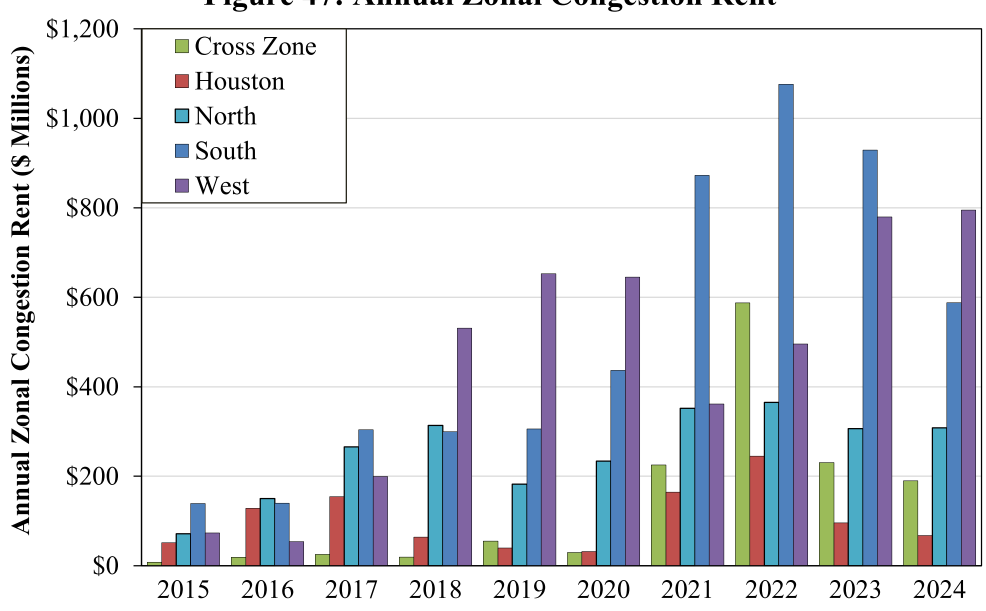
Other upcoming policy changes also highlight the growing need for load zone pricing that reflects the actual cost of serving load within each zone. For example, with the approval of NPRR 1188 in November 2024, Controllable Load Resources (CLRs) will be shifted from zonal to nodal pricing, removing their nodal prices from the Load Zone (LZ) price calculations. This policy will incentivize CLRs to site at lower priced nodes, thus removing those lower priced nodes from the calculation of the load zone price, resulting in higher load zone prices. For some customers, such as inflexible oil and gas load in the Permian Basin, higher prices resulting from load zone reconfiguration will more accurately reflect their true cost of service. However, for consumers in the Panhandle who lack the flexibility to qualify as CLRs, the exodus of more flexible load from the zone will only widen the gap between the prices they are charged and the actual cost of serving them. This highlights the need for careful design of load zones to ensure that price signals align with cost causation across all types of customers.

This section introduces a methodology for re-defining the load zones according to geographic proximity and historical nodal prices. We then present an analysis of how the implementation of these updated load zones would impact congestion management and zonal pricing outcomes.

1. Congestion Impact

The four current load zones within ERCOT were established in 2003 and comprise the North, West, South, and Houston load zones. These load zones no longer effectively represent the dynamics of Texas's electricity market and result in high rates of intra-zonal congestion, particularly in the South and West Zones, as shown in Figure 47.

Figure 47: Annual Zonal Congestion Rent

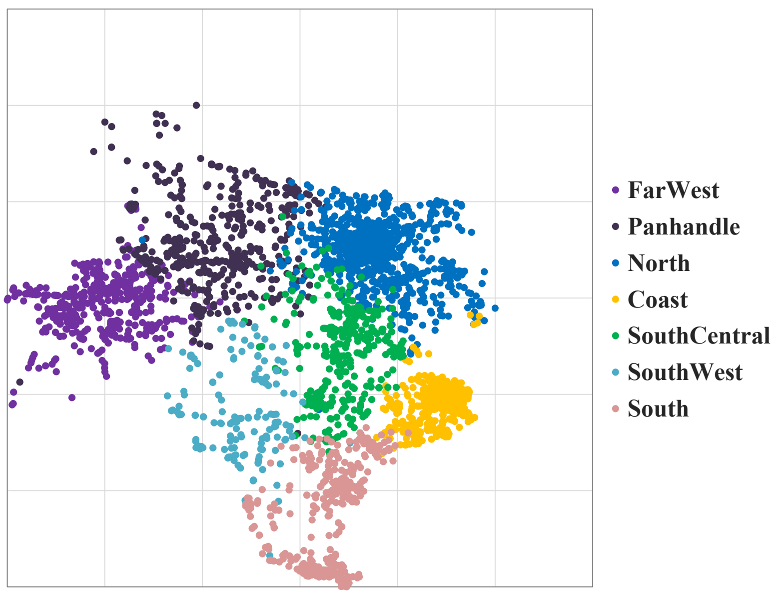


This intra-zonal congestion represents a growing difference in the cost of service within the load zones that should be reflected in the prices paid by load. To address this issue, the IMM recommended in the 2020 SOM report that ERCOT should reconfigure the load zones to better reflect the topology of the network. Next, we discuss a methodology for defining the boundaries of the load zones based on geographic coordinates and historical pricing outcomes.

2. Methodology for Defining Load Zones

The methodology groups substation-level load nodes into new load zones based on geographic coordinates and historical price data.[38] These metrics were chosen to define load zones according to proximity and congestion conditions. Our analysis evaluated configurations of six, seven, and eight load zones, spanning from January 2021 to December 2024. Figure 48 illustrates the resulting distribution of load nodes within the proposed seven-load-zone configuration.

Figure 48: Geographic Distribution of Substations for the 7-Load-Zone Configuration



To evaluate the improvement in zonal pricing achieved by this updated load zone configuration, we consider the resulting decrease in intra-zonal congestion rent, as shown in Table 5. This data indicates that such a reconfiguration would result in a significant reduction in intra-zonal congestion compared to the current load zone map. This reconfiguration also produces more congestion rent between zones, the result of price disparities that efficiently reflect the geographic differences in the cost of serving loads in different parts of the grid.

38 The methodology uses k-means clustering refers to a machine learning algorithm used to group data into clusters based on their similarities. This algorithm incorporates geographic proximity, congestion data, and a specified number of load zones to arrive at a grouping of substations into a new set of load zones.

Table 5: Real-Time Congestion Rent ($MM) for the 7-Load-Zone Configuration

	Cross Zone	Coast	North	South	SouthCentral	SouthWest	FarWest	Panhandle
				$ Millions				
2021	$821.6	$304.5	$208.8	$214.9	$90.9	$78.6	$90.3	$223.7
2022	$1,454.4	$387.8	$185.2	$268.3	$49.3	$122.2	$119.3	$187.7
2023	$1,176.2	$230.8	$151.5	$171.3	$122.9	$142.2	$59.9	$280.3
2024	$853.2	$100.5	$105.4	$215.6	$180.8	$85.6	$164.7	$254.5

For more detail on these disparities, Figure 49 and Figure 50 compare pricing for the West and South Load Zones to the prices corresponding to our proposed configuration of seven load zones.

Figure 49: Comparison of LZ West Prices: Current vs. New 7-Load-Zone Configuration

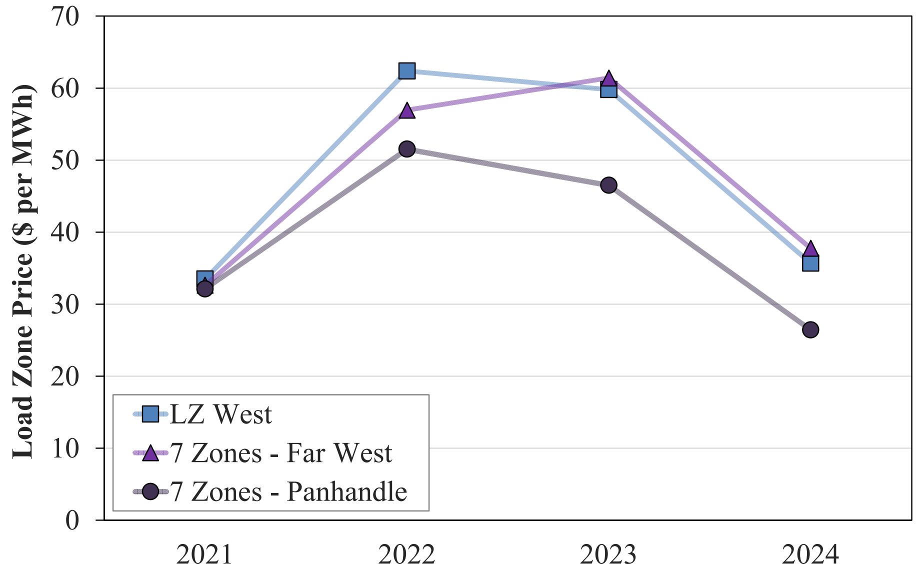


Figure 50: Comparison of LZ South Prices: Current vs. New 7-Load-Zone Configuration

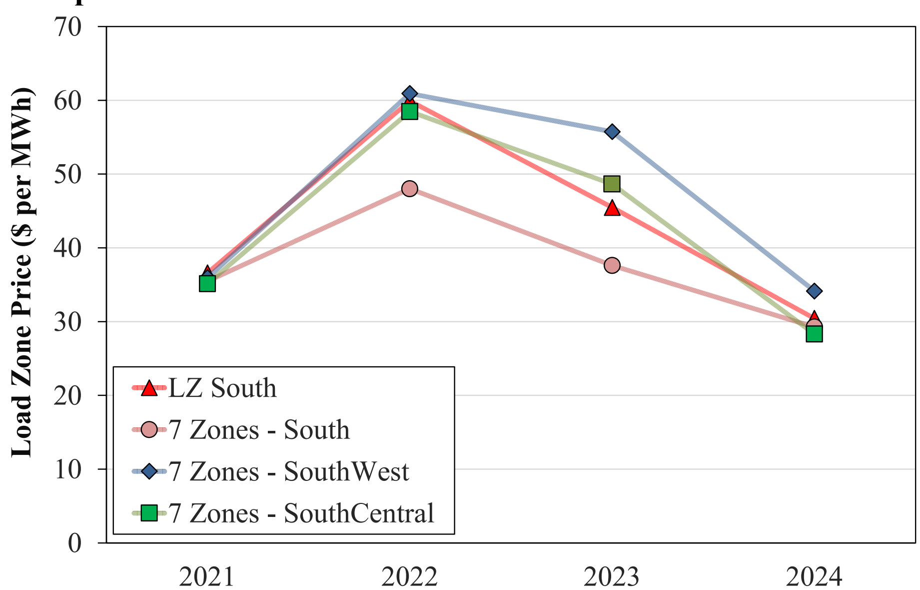


Figure 49 shows a significant difference in pricing between the Panhandle and Far West, which approximately corresponds to the Permian Basin. Similarly, Figure 50 shows large differences in pricing between a new "7-South" load zone that goes from Corpus Christi to the Rio Grande Valley and the new SouthWest and SouthCentral zones. These large differences in pricing between regions included in the same load zone indicate that the current load zones are obscuring large differences in the cost of serving load within the same zone.

This analysis supports our recommendation that ERCOT develop a process for re-defining the load zones on some basis so that they better reflect the current congestion conditions on the grid. Given that CRR auctions are conducted as far as three years into the future, any changes to the definitions of the load zones would likely have to be set at least three years in advance so that market participants have sufficient time to factor any changes into their CRR positions. Additional detail related to this analysis can be found in our presentation to the Congestion Management Working Group (CMWG) in July 2024.[39]

[39] https://www.ercot.com/files/docs/2024/07/10/Updating-ERCOT-Load-Zones-CMWG-July-2024.pptx

V. Market Operations

Ideally, markets should procure and utilize all of the resources necessary to reliably operate the system. In reality, the market schedules and instructions are often supplemented by out-of-market actions by the operators to address operational issues. Out-of-market actions are undesirable because they interfere with the price signals that drive efficient short-term behavior and long-term investment decisions in a competitive electricity market. These actions can also lead to cost shifts between market participants, reduce transparency, and complicate market settlements. While sometimes necessary for reliability, frequent reliance on out-of-market actions suggests a misalignment between the market requirements and the operational needs of the system. This chapter focuses on these types of out-of-market operator actions.

A. Reliability Unit Commitments

1. Unit Commitment under the Multi-Settlement Market

Shortfalls in market-procured capacity can arise from how generators participate in the market. The majority of generators in ERCOT decide on their own whether to start up, a practice known as self-commitment seen in Figure 30. This approach contrasts with other Independent System Operators (ISOs) or Regional Transmission Organizations (RTOs) where a much larger share of the generation is scheduled through the day-ahead markets.

ERCOT's day-ahead market is financially binding and does not result in physical obligations in real time. In other words, a generator scheduled in the day-ahead market is not required to generate electricity in real-time – it has the option of not running and buying back the day-head schedule at the real-time price. Conversely, if it delivers more than its day-ahead award, it is paid the real-time price for the generation in excess of its day-ahead schedule. This multi-settlement system aligns participants economic incentives to be available with the operational needs of the system. Nonetheless, if ERCOT projects that insufficient generation will be available in real-time, it may issue an out-of-market commitment instruction through the Reliability Unit Commitment (RUC) process that is described below.

2. RUC Fundamentals

ERCOT can commit additional generators that were not either self-committed or scheduled day ahead through the RUC process. RUC commitments can occur either in the day-ahead timeframe, known as Day-Ahead RUC (DRUC) or closer to real time through Hourly RUC (HRUC). The vast majority of RUC instructions come out of the HRUC process. For resources that submitted a valid three-part offer in the day-ahead market, RUC uses these offers. For all other resources, RUC uses either verifiable cost data or, in the absence of verifiable cost data, generic cost data associated with different classes of resources.

RUC commitments increase the supply of generation in the market, placing downward pressure on prices. To reduce this price distortion, ERCOT uses the Reliability Deployment Price Adder (RDPA), which adjusts real-time prices upward to account for the additional supply injected by RUC, thereby preserving shortage signals that would have existed without the out-of-market commitment. It also applies an offer floor of $250 per MWh for RUC-committed units. By setting a high minimum-offer price for these resources, ERCOT reduces the likelihood that they will be economically dispatched or set the market clearing price, limiting their direct influence on real-time prices. Together, these tools help limit the extent to which RUC suppresses prices.

Operators issue RUC commitments to meet forecasted system-wide demand or to manage congestion. In the latter case, specific units may be required to serve load in transmission-constrained areas, to provide counterflow on a constraint, or support local reliability. The criteria for making RUC commitments should be transparent and grounded in objective reliability risks rather than driven by an arbitrarily conservative operational approach. Risk-based standards help ensure that the RUC process is used only when necessary, which supports market efficiency, and to maintain stakeholder confidence in ERCOT's operational decisions.

3. Make-Whole Payments and Clawbacks

As discussed in the previous section, generator operating costs are incorporated into the RUC process, either through three-part offers, verifiable costs, or generic costs inputs by resource type. When a generator is committed through RUC based on these costs, ERCOT uses the cost data to determine whether the unit is entitled to a make-whole payment. These payments ensure that RUC-committed units are not financially harmed when their market revenues fall short of their costs. Conversely, if a RUC-committed unit earns more revenue than its costs, ERCOT may apply a clawback to recover some or all of the excess revenues, depending on whether a valid three-part offer was submitted in the day-ahead market.

The cost of make-whole payments is allocated to two groups. First, Qualified Scheduling Entities (QSEs) that do not provide enough capacity to cover their real-time obligations are considered capacity short and bear a portion of the cost. Second, all QSEs share the remaining costs on a load-ratio-share basis. Suppliers also have the option to opt-out of both the make-whole payment and any associated clawback, which effectively means self-scheduling the unit and accepting full exposure to market outcomes. This approach gives suppliers the flexibility to manage their own risk while helping ERCOT maintain system reliability.

Prior to 2024, RUC-committed resources that had submitted valid day-ahead offers were subject to only a 50% clawback of revenues above their costs. This partial clawback created a financial incentive for some units to avoid self-committing, even when they were likely to be economic,[40]

[40] It is notable that there is no requirement that the day-ahead market energy offer that triggers the reduced claw-back percentage be feasible, i.e., able to be awarded by the day-ahead market engine based on resource temporal constraints.

in order to benefit from the opportunity to recover all costs and retain half of any upside through RUC.[41] In response to concerns that this undermined efficient market behavior, consumer stakeholder representatives filed NPRR 1172, *Fuel Adder Definition, Mitigated Offer Caps, and RUC Clawback*, in April 2023.[42] The proposal called for 100 percent clawback of revenues above cost for economic resources that were RUC-committed after submitting day-ahead offers. The PUCT approved NPRR 1172 and it took effect on March 1, 2024.

Since the implementation of this rule change, patterns in RUC clawback and make-whole payments have shifted. As shown in Table 6, clawback payments increased while make-whole payments declined in 2024. This increase in clawbacks was largely driven by a high number of unit hours in March and April of 2024 during which RUC-committed resources did not opt out of RUC settlement. We discuss these behaviors and additional RUC trends next.

Table 6: RUC Settlement Quantities, 2020-2024

	Claw-Back ($MM)	Make-Whole ($MM)
2020	$0.48	$0.40
2021	$3.09	$5.38
2022	$23.74	$42.78
2023	$3.07	$3.63
2024	$7.41	$2.69

4. RUC Trends

We now examine several recent trends related to the frequency of RUCs, the reasons RUCs were issued, and generator opt-out behavior. These trends are summarized in Table 7.

Table 7: Reasons for RUC, 2020-2024

Year	# of RUC-Resource hours		% of RUC-Resource hours	
	Congestion	Capacity	Congestion	Capacity
2020	224	-	100%	-
2021	810	3,242	20%	80%
2022	1079	7,166	13%	87%
2023	295	2,439	11%	89%
2024	738	1,237	37%	63%

41 The IMM recommended that ERCOT eliminate the 50% claw-back for day-ahead offers and implement a 100% claw-back for economic RUC resources in its 2022 State of the Market Report (see Recommendation 2022-2) and filed comments supporting NPRR 1172.

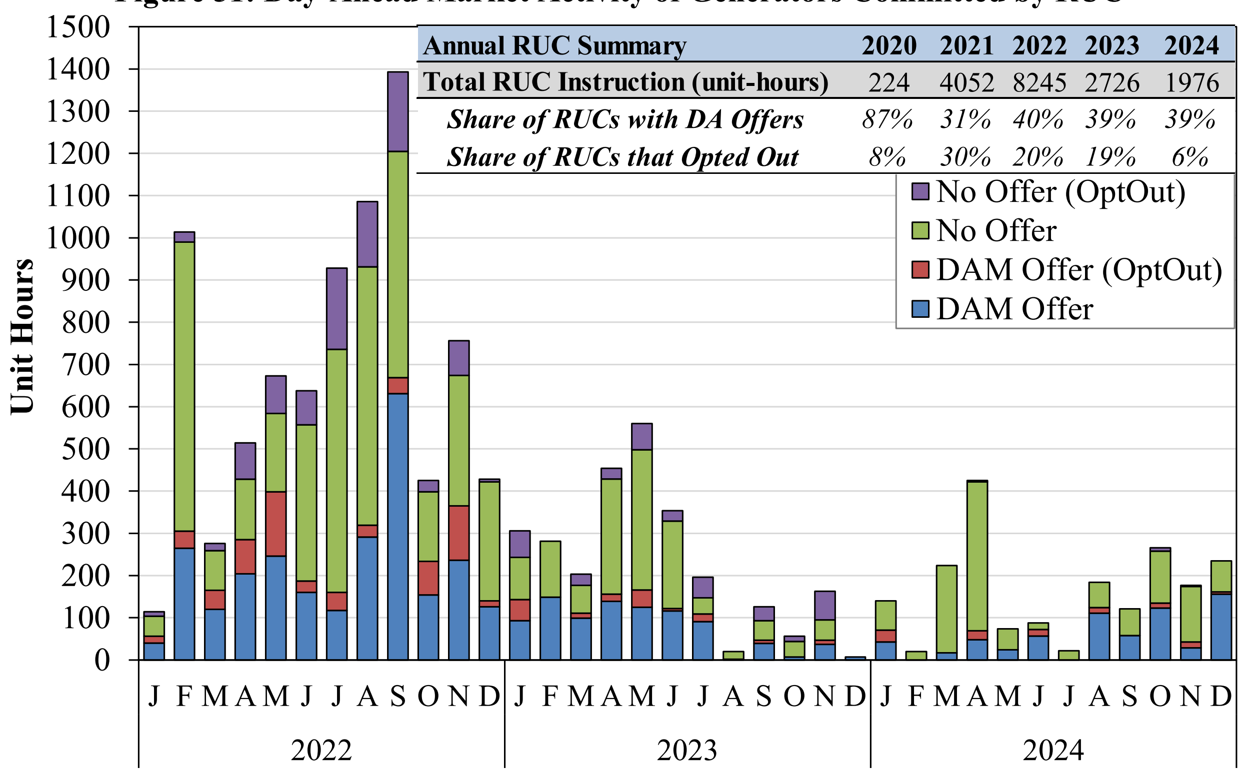
42 NPRR 1172, *Fuel Adder Definition, Mitigated offer Caps, and RUC Clawback*, available at: https://www.ercot.com/mktrules/issues/NPRR1172.

Prior to 2021, ERCOT introduced process improvements that significantly reduced the frequency of RUC commitments and most RUCs before this time were issued to manage transmission congestion. This pattern shifted abruptly in June 2021, when ERCOT adopted a more conservative operational approach. Under this new posture, ERCOT began committing additional generation resources and doing so earlier in the operating day. As a result, RUC activity increased sharply from mid-2021 through mid-2023, with most of the new commitments driven by system-wide capacity needs rather than local congestion management.

Figure 51 shows monthly RUC activity over the past three years and distinguishes between units that submitted day-ahead offers and those that opted out of RUC settlement. From 2021 through the first half of 2023, RUC activity was elevated as ERCOT maintained a conservative operational approach. This was followed by a notable decline in RUC commitments during the second half of 2023. The decline was driven, at least in part, by higher energy and operating reserve prices following the implementation of ERCOT Contingency Reserve Service (ECRS), which encouraged more self-commitments.

In 2024, RUC activity remained below earlier levels but increased slightly compared to late 2023. This rebound coincided with the addition of substantial new solar capacity, which contributed to lower prices and reduced self-commitment from thermal units. In some cases, over-forecasting of solar generation may have led to concerns about real-time capacity, prompting ERCOT to issue more RUC commitments.

Figure 51: Day-Ahead Market Activity of Generators Committed by RUC



Annual RUC Summary	2020	2021	2022	2023	2024
Total RUC Instruction (unit-hours)	224	4052	8245	2726	1976
Share of RUCs with DA Offers	*87%*	*31%*	*40%*	*39%*	*39%*
Share of RUCs that Opted Out	*8%*	*30%*	*20%*	*19%*	*6%*

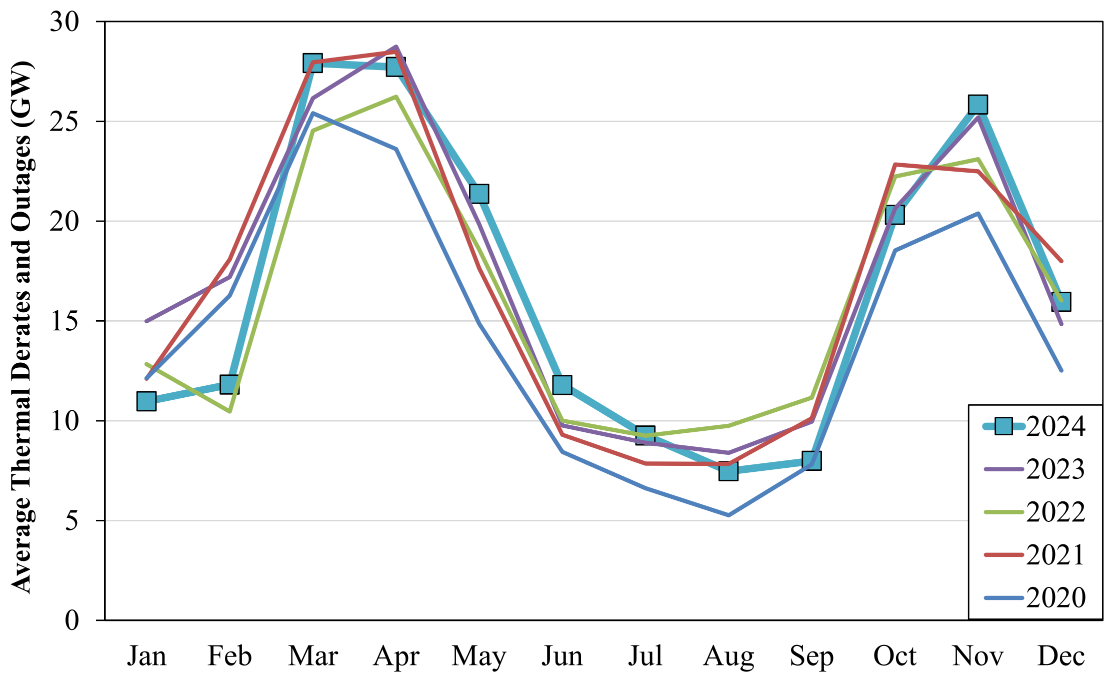
In 2020, 87% of RUC-committed units had submitted valid three-part offers in the day-ahead market. Starting in 2021, this percentage dropped significantly to between 31% and 40%, where it remained through 2024. This shift coincided with increased use of RUC following ERCOT's adoption of a more conservative operational posture. During the early part of this period, a higher share of RUC-committed units also opted out of RUC settlement, which reduced the number of units eligible for make-whole payments or subject to claw-backs. By 2024, only 11% of RUC-committed units opted out of settlement, returning closer to pre-2021 levels. This shift is likely another consequence of NPRR 1172.

B. Thermal Generation Outages and Deratings

At any given time, some portion of ERCOT's generation is unavailable because of outages and deratings. Derated capacity is the difference between the registered summer maximum capacity of a resource and its actual capability. It is common for generating capacity to be partially derated because the resource cannot achieve its installed capacity level due to technical or environmental factors (e.g., equipment failures or ambient temperatures).

Outages and deratings of thermal power plants are especially important because they can affect reliability during periods when other resources are limited. As ERCOT has become more reliant on wind and solar, overall generating capacity has grown more sensitive to weather conditions. During times of high demand with low renewable output, thermal units often become essential for meeting system needs. However, thermal generators tend to schedule more of their outages in the spring and fall, when demand and prices are typically lower. Figure 52 illustrates this seasonal pattern.

Figure 52: Thermal Hourly Average Outages and Derates by Month



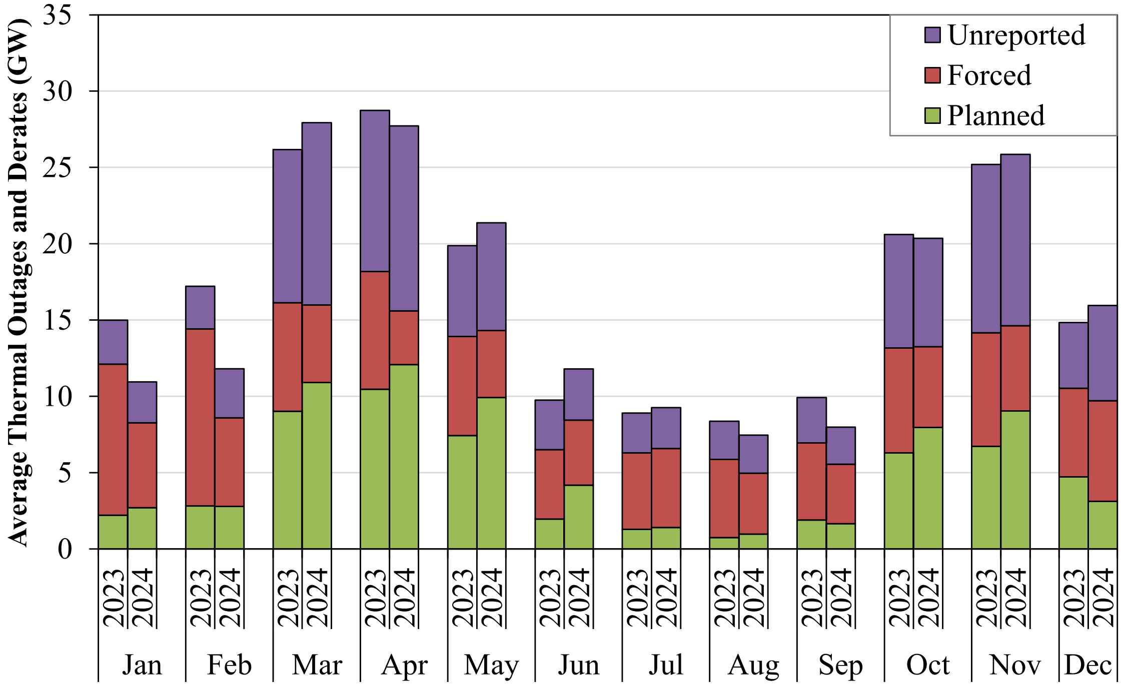
Outages and derates introduce uncertainty for ERCOT operators and market participants by complicating real-time assessments of supply and pricing. To help system operators to manage this dynamic, generators are expected to schedule planned outages in advance, giving visibility into unavailable capacity. However, planned outages account for only part of the total. Many outages are forced, resulting from unexpected failures that take units offline. While most forced outages are eventually reported, a large number remain unreported, making it harder for ERCOT to plan and operate the system reliably. Strengthening outage and derate reporting requirements would help improve system coordination and transparency.

Figure 53 presents monthly totals of planned, forced, and unreported outages and derates of thermal resources for 2023 and 2024. For records in the outage scheduler occurring less than 30 days, the notification for an outage greater than 7 days to the start of the outage was considered planned, a report less than or equal to 7 days prior was considered forced. Patterns in 2024 were similar to those in 2023, with planned and unreported outages lowest during the summer, when energy is most valuable, and highest in the spring and fall, when system load and net load are typically lower. Forced outages remained relatively stable across the year. NPRR 1084, Improvements to Reporting of Resource Outages, Derates, and Startup Loading Failures, was implemented at the end of 2022 to improve outage reporting practices.[43] However,

Figure 53 indicates that a large share of outages and derates in 2024 were still not reported in the outage scheduler.

43 https://www.ercot.com/mktrules/issues/NPRR1084

Figure 53: Planned, Forced, and Unreported Outages and Derates



C. QSE Operation Planning

The Current Operating Plan (COP) is the mechanism used by QSEs to communicate the expected status of the QSE's resources to ERCOT. COPs are updated on an ongoing basis by QSEs for each operating hour. The RUC process uses the schedules reported in the COP to see which resources are planning to be running each hour of the operating day. The schedule of commitments in COP along with forecasts for renewable generation are compared against forecasted load to determine if out of market commitments are necessary to manage a system-wide supply shortage or transmission constraint. Resources shown as offline in their COP are eligible for commitment through RUC subject to start-time constraints. Thus, the accuracy of COP information greatly influences ERCOT's ability to effectively commit resources through the RUC process.

To summarize the accuracy of COP statuses in situations where RUC may be needed for capacity, we considered all intervals where online reserves were less than or equal to 6,500 MW. We then compared the real-time status of all resources in SCED against their COP for the last submitted COP that would have been seen before a decision had to be made about committing a unit given their start time. For example, if a resource has a start time of six hours, we compared the real-time status of that resource against the status reported in its COP from six hours prior. Figure 54 summarizes the magnitude of disparity between real-time and COP statuses using this methodology is shown in.

Figure 54: Difference in Real-Time vs COP Capacity, PRC ≤ 6,500 MW

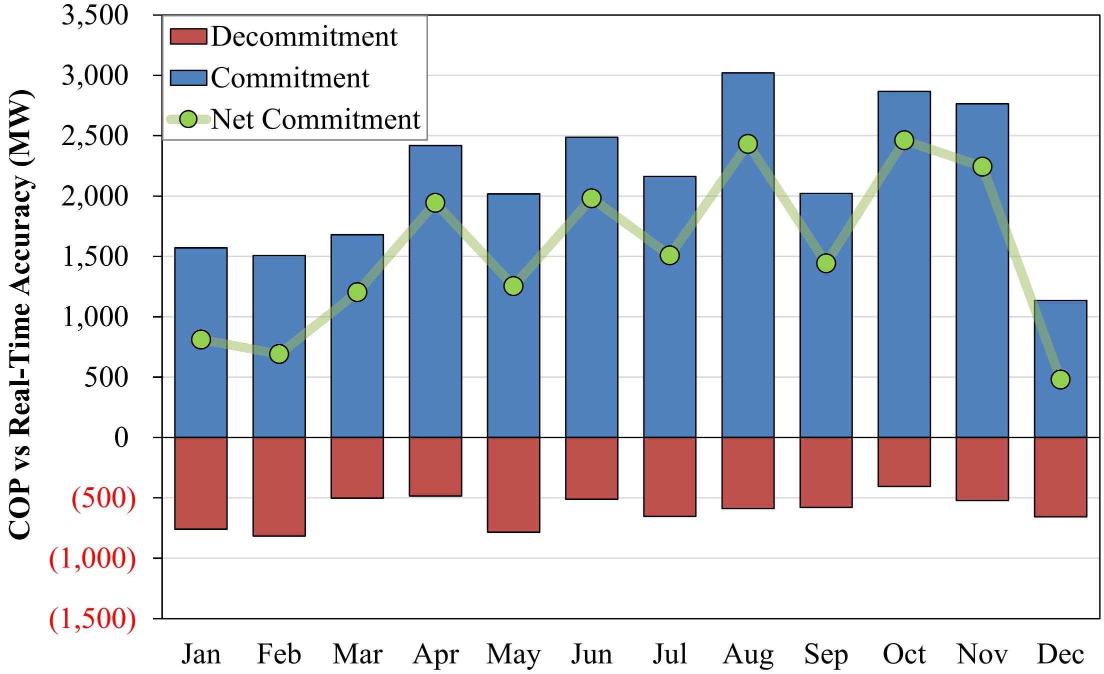


Figure 54 shows that for intervals where online reserves were less than or equal to 6,500 MW there is on average approximately 1,300 MW more capacity online in real-time than was reported in the last applicable COP, where the net difference in capacity between real-time and COP is plotted in green. The blue bars refer to capacity that was online in real-time when it was scheduled to be off in COP, and the red bars refer to capacity that was offline in real-time when it was scheduled to be on in COP.

The most noteworthy aspect of this data is that over 90% percent of the inaccuracy of net committed capacity in COP can be attributed to resources with a start time of one hour or less. Resources with longer start times tend to have more accurate COPs corresponding to intervals with relatively low levels of reserves in real-time. Thus, the magnitude of COP inaccuracy is less problematic than it may appear at first, because there is less risk in waiting until closer to real-time to commit units with shorter start times. Even after the last regular RUC run before real-time, operators can manually commit these short start-time units if necessary.

D. Firm Fuel Supply Service

A new Firm Fuel Supply Service (FFSS) was approved and implemented in 2022, which pays a subset of dual-fuel generators to purchase fuel to be stored on site.[44] As of July 1, 2023, FFSS was expanded to also include certain gas-fired generation resources with owned natural gas

---

44 *See* https://www.ercot.com/services/programs/firmfuelsupply; NPRR 1120, *Create Firm Fuel Supply Service*, available at: https://www.ercot.com/mktrules/issues/NPRR1120.

stored offsite and accompanied by firm transportation and storage agreements.[45] Implementation of FFSS was part of the PUCT's Phase I Market Design effort and in response to Texas Senate Bill 3, 87th Session.

ERCOT has now issued three RFPs for FFSS, each with an obligation period beginning on November 15 and ending on March 15.[46] In 2024, FFSS was deployed across five consecutive days from January 13-17. Over that time, 10 different FFSS Resources (FFSSRs) were deployed for a maximum of 916 MW, as shown in Table 8.

Table 8: Firm Fuel Supply Service Deployments

Day	Maximum Aggregate FFSS Deployment (MW)	Average RT Price	Operating Day Online Reserves Minimum
1/13/2024	80	$3.66	17,358
1/14/2024	726	$70.78	12,464
1/15/2024	916	$88.02	8,397
1/16/2024	916	$148.71	5,414
1/17/2024	150	$27.49	15,054

ERCOT's FFSS Deployment Report for this event states that the decision to deploy FFSS was based on information about potential gas supply restrictions that could affect generation resources. However, if such restrictions did occur, they did not lead to a noticeable decline in operating reserves or a rise in real-time prices, as shown in Table 8. Only January 16 saw conditions tight enough to produce a significant ORDC price, reaching $88 per MWh for one SCED interval. These outcomes raise questions about whether FFSS deployments were necessary on most of the days included in this event.

The procurement and deployment of FFSS costs ERCOT consumers tens of millions of dollars per year. One factor that contributes to inefficient market outcomes and excess cost is that FFSS resources are also eligible for make-whole payments if their market revenues during a deployment are less than the cost of replacing their spent fuel. This make-whole payment diminishes the incentive for FFSS resources to offer at prices reflecting the marginal cost of replacement fuel. If FFSS resources were required to offer according to the marginal cost of fuel replacements, those resources would be dispatched at lower levels, resulting in lower make-whole payments.

Further, real-time price is distorted with FFSS deployments when the cost of the deployed assets are not accounted for in price formation. This has a price-suppressing effect for all other supply

45 NPRR 1169, Expansion of Generation Resources Qualified to Provide Firm Fuel Supply Service in Phase 2 of the Service, available at: https://www.ercot.com/mktrules/issues/NPRR1169.

46 *Wholesale Electric Market Design Implementation*, Project No. 53298, ERCOT Letter Regarding FFSS Phase I Procurement Results (Sept. 27, 2022). ERCOT Report of the Second Procurement of the Reliability Product, Firm Fuel Supply Service (FFSS) (Sept. 21, 2023).

in the real-time market. Another factor is that the aggregate high sustained limit (HSL) of deployed FFSSRs is not included as online reserves in the calculation of the ORDC price adder, which results in higher shortage prices. Excluding this capacity from the calculation of the ORDC resulted in almost $7 million of additional cost for this event.

To address these issues, we proposed the following improvements that are consolidated in Recommendation 2023-4:

- ERCOT should develop clear procedures based on reliability metrics for deploying FFSS. For example, forecasted generation shortfalls or unresolvable transmission constraints caused by disruptions in the gas supply;
- Require all deployed FFSSRs to offer according to the marginal cost of replacing their spent fuel to minimize the need for make-whole payments and reduce price suppression; and
- Include FFSS capacity in the calculation of the ORDC. Removing the entire HSL of FFSSRs from RTOLCAP unreasonably increases the cost of shortage pricing relative to available capacity.

VI. Resource Adequacy

A. Introduction

Ensuring resource adequacy is fundamental to the reliability and stability of the electricity market. Resource adequacy refers to the availability of sufficient generation and demand-side resources to meet expected electricity demand and ancillary services under normal and extreme conditions. A well-functioning market must send clear price signals to incentivize investment in new generation, maintenance of existing resources, and demand-side participation. Without these signals, the market risks underinvestment in critical infrastructure, leading to reliability challenges and potential supply shortages.

Generators assess resource adequacy to identify investment opportunities and the potential for higher revenues during shortage conditions. Load-Serving Entities (LSEs) and large consumers monitor adequacy to anticipate price volatility and plan strategies for cost management, such as demand response. A well-functioning market provides price signals for all participants to plan effectively, adapt to changing conditions, and ensure long-term system reliability.

The following concepts are important to understand regarding revenue sufficiency and investment in new generation:

Cost of New Entry (CONE): CONE represents the estimated fixed expense of building and operating a new power plant. Investors evaluate whether expected future market revenues will be sufficient to justify these costs before committing to new projects.

- **Shortage Pricing:** In electricity markets, prices rise during periods of tight supply to reflect the increased value of available generation. These price spikes create opportunities for generators to recover fixed costs and incentivize new investment. In ERCOT's energy-only market, shortage pricing serves as the primary mechanism for driving revenue and signaling investment decisions.
- **Peaker Net Margin (PNM):** PNM estimates the annual net revenue a peaking unit could have earned based on observed energy and ancillary service prices. Comparing PNM to CONE helps market participants determine whether revenues are sufficient to support new generation or if additional incentives may be needed to maintain resource adequacy.

1. Key Reports

ERCOT communicates expectations regarding future resource adequacy through several reports that provide market participants with insight into future conditions in different timeframes.

- The Monthly Outlook for Resource Adequacy (MORA) offers a short-term assessment of expected supply and demand conditions, highlighting potential risks in the coming months.

- The Capacity, Demand, and Reserves (CDR) report provides a 5-year forecast of load growth and generation capacity to help participants evaluate future resource adequacy.
- Complementing these reports, the Long-Term Load Forecast (LTLF) projects demand trends over a period of up to ten years, offering a broader perspective on future needs.

Together, these reports help market participants anticipate challenges, identify investment opportunities, and plan accordingly.

2. Resource Adequacy through Markets

The economic signals provided by the wholesale electricity markets will facilitate long-term investment and retirement decisions that maintain an economic level of capacity that is consistent with these signals. In general, there are three primary approaches to achieve adequate resources through competitive wholesale electricity markets:

1. <u>Energy-only market</u> – this market relies primarily on expected shortage revenues in the energy and ancillary services markets to motivate investment.
    - Pros: Provides strong performance and availability incentives. Is closely aligned with reliability.
    - Cons: Capacity levels may be less than needed to satisfy a particular reliability target, such as the "1-in-10" standard adopted by most Regional Transmission Organizations (RTOs).[47] Can produce highly volatile year-to-year costs and revenues that can be hedged by contracts.
2. <u>Capacity market</u> – Designed to procure a sufficient quantity of capacity to satisfy a specified reliability standard
    - Pros: Predictably generates the revenues needed, together with the energy and ancillary services net revenues, to compensate investors to build generation that will maintain this level of capacity, i.e., to cover new resources' net CONE.
    - Cons: Requires more complicated rules related to accreditation of generation and load resources. Is generally less directly aligned with specific operational reliability needs, for example Non-Spin Reserve Service (NSRS) or frequency response service. Capacity constructs procure generic capacity.
3. <u>Capacity requirements</u> – Some markets require LSE's to self supply or procure capacity to satisfy a specified capacity requirement. This is effectively a decentralized capacity market that operates bilaterally.
    - Pros: Increases the likelihood of satisfying the specified reliability standard.
    - Cons: Prices may not be efficient or competitive, which could raise costs compared to a centralized capacity market procurement.

47 The 1-in-10 standard is the capacity needed to expect to shed load in one event each 10 years.

All market-based market proposals would fall within one of these three approaches. Each approach includes details that can be adjusted to achieve specific objectives. For example, capacity markets include many choices of design, including: (1) the procurement timeframe (prompt auction that run in the months before the planning year vs. forward auctions that run up to 3 years ahead), (2) resource accreditation rules, (3) capacity demand curve estimation, and (4) market power mitigation measures.

In an energy-only market, the shortage pricing will be the result of operating reserve demand curves (ORDCs) that will set prices when the market does not have sufficient resources to satisfy the full market requirements for energy and ancillary services. If the market is not sustaining sufficient resources, the economic signals can be strengthened by increasing the aggregate value implied by the ORDC. The Commission implemented such changes at the start of 2022 and they have been extremely effective in increasing the markets' shortage revenues. Note that spot market shortage pricing is an important element of market design even in instances where there is a capacity construct as well. Shortage pricing values additional supply at the marginal value it contributes to reliability and signals both performance and new investment. When a system is capacity-short (resource inadequate), spot market revenue will increase through more frequent and severe shortage pricing that can accompany shortage signals from the capacity construct.

Ultimately, the ORDC implies a "value of lost load" (VOLL), which is the value of avoiding load shedding. One key issue with satisfying the typical 1-in-10 reliability standard adopted by most of the RTOs throughout the country is that this standard implies a VOLL in excess of $200,000 per MWh. This explains why most RTOs have had to rely on capacity markets to supplement the revenues from the energy and AS markets to satisfy this standard. This also reveals why it is difficult to satisfy such a standard in an energy only market.

Other issues include the potential need to consider resource adequacy requirements for import-constrained zones, potential market power for new and existing resources when the market is relatively capacity-short, and market power mitigation and backstop procurement (in the load obligation model) when load and supply entities are not able to agree on a competitive price for existing or new build capacity.

3. Focus Areas of this Chapter

In the following sections, we examine the key factors influencing investment and resource adequacy in ERCOT, including:

- Analyzing the net revenues earned by various generation technologies in different locations, offering insight into how market conditions affected generator profitability;
- Evaluating CONE and market revenues to assess whether recent market revenues have been sufficient to support new investment;
- Discussing the reliability standard introduced in 2024 and its implications within ERCOT's energy-only market structure;

- Reviewing the primary reports that communicate the load and generation trends that determine ERCOT's resource adequacy, highlighting the limitations of these reports; and
- Reviewing the events that have resulted in the current load forecast process and offering our insights into the most recent data published in April 2025.

B. Net Revenue Analysis

We calculate net revenue by subtracting a generating unit's variable production costs from its total potential revenue. In other words, net revenue represents the earnings available beyond short-run operating costs to recover fixed and capital expenses, including a return on investment. Net revenue is the key determinant of the incentive to invest because it is the earnings available to recover fixed and capital expenses, including a return on investment, after short-run operating costs are covered. In ERCOT's energy-only market, net revenues from the energy and ancillary services markets serve as the primary economic signals guiding investment and retirement decisions for generation resources. While revenues may also come from the day-ahead market or forward bilateral contracts, these ultimately reflect expectations of real-time energy and ancillary service prices. Although the net revenues presented in this report are based on historical prices, investment decisions are typically driven by expectations of future market conditions, including the potential for shortage pricing.

1. Peaker Net Margin and the ORDC

The peaker net margin (PNM) and shortage pricing mechanisms like the ORDC play a crucial role in shaping net revenue and investment signals in the electricity market. PNM estimates the annual net revenue a peaking unit could earn from energy and ancillary service markets, serving as a benchmark for evaluating whether market conditions support new investment. If PNM approaches or exceeds the CONE, it suggests that market revenues are sufficient to support new generation for that year. The ORDC reinforces this by raising energy prices when operating reserves fall below predefined thresholds, ensuring that generators are compensated for providing reliability during shortage conditions. The Reliability Deployment Price Adder (RDPA) also raises energy prices by accounting for grid operator actions made to maintain reliability that impact the market.

2. Net Revenue by Location

Figure 55 shows the net revenues at different locations for a variety of new generators. Because natural gas prices can vary widely based on location, the revenues for natural gas units are shown for the Houston zone (reflecting Katy hub prices) and the West zone (reflecting Waha hub prices).

Figure 55: Net Revenues by Location, 2024

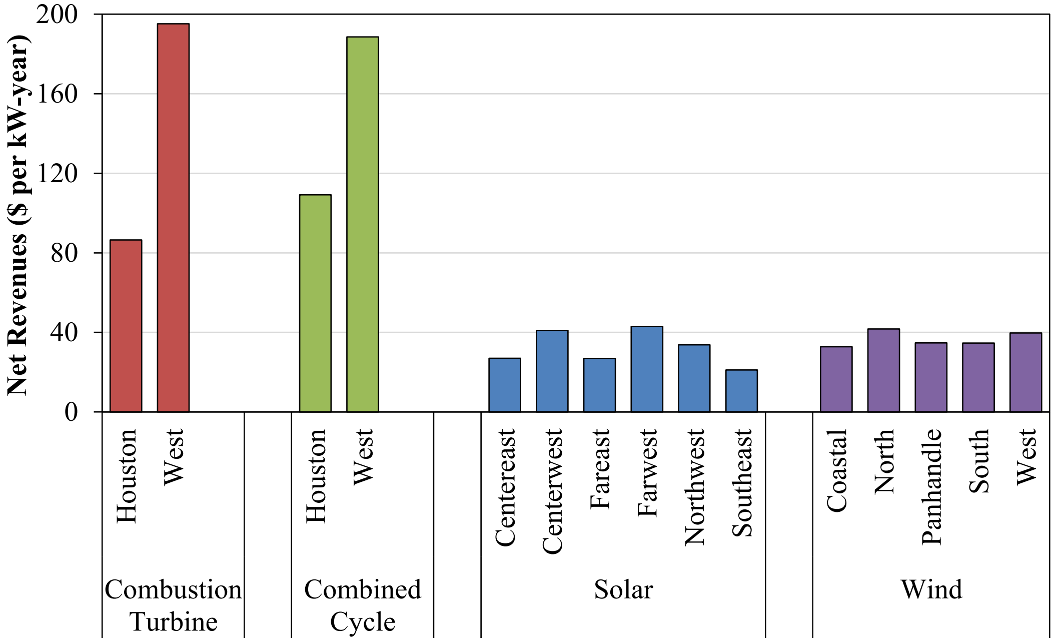


Figure 55 shows a wide gap between the net revenues in the West and Houston. Historically, high natural gas production in the Permian Basin and limited export capability have resulted in low gas prices at the Waha location and, as a result, much higher net revenues for gas resources in this area. The price gap between the two hubs widened in 2024, driven by transmission upgrades and maintenance in the West zone that substantially increased congestion costs in the region. As work on these projects concludes, we expect the gap between these two hubs to narrow again.

Figure 55 also shows the net revenues for wind and solar generation at multiple locations. The profitability of these resources is primarily driven by the amount of the local wind or solar penetration and the market prices during periods of high output. In 2024, net revenues for wind and solar were lower than those of gas-fired technologies across all areas. Additionally, the locational spread in net revenues across IRRs was smaller in 2024, due in part to more uniform weather patterns and relatively consistent congestion patterns across regions, which reduced the revenue advantages typically seen in higher-performing locations.

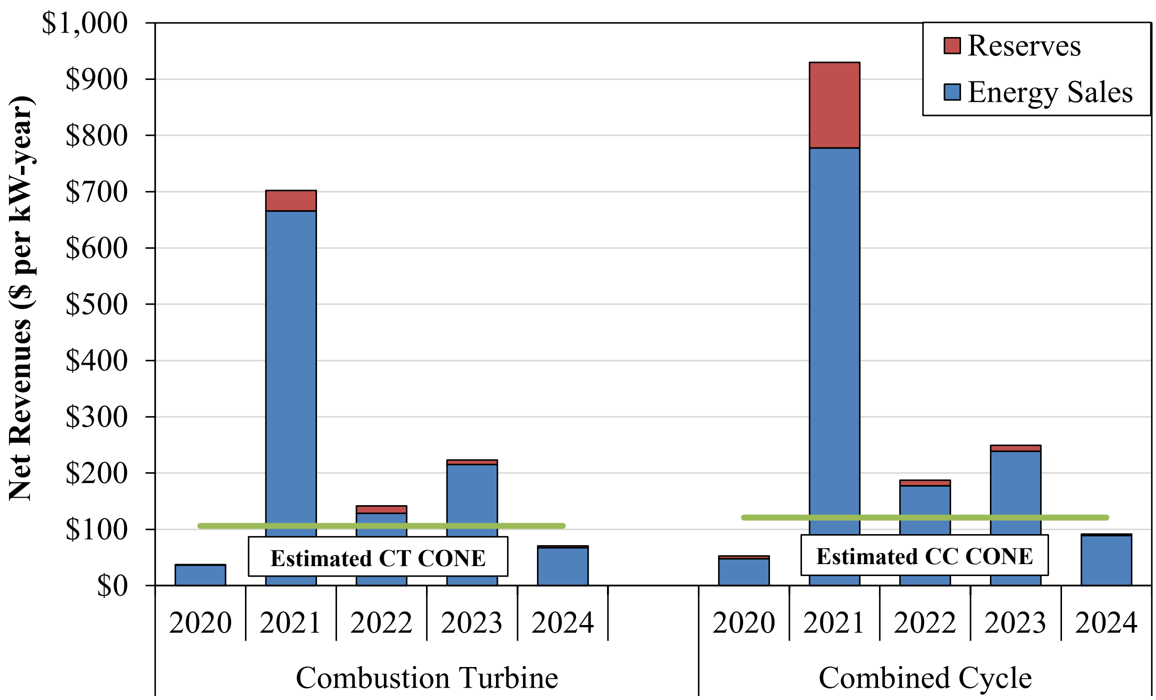
C. Cost of New Entry and Net Revenues

CONE represents the minimum annual revenue a new generator, typically a gas-fired unit, must earn to recover both its capital and fixed operating costs over its expected lifetime. The reference technology is chosen to reflect the technology and configuration that is most likely to be built by a merchant developer in response to market price signals. Functionally, CONE serves as a levelized cost of electricity (LCOE) benchmark, capturing the amortized revenue requirement per kW for a generator to be economically viable. CONE is often framed in terms of gas generators, as they frequently serve as the marginal units in ERCOT's energy market and therefore provide a useful reference point for assessing the adequacy of market price signals to support new investment.

In practice, generator earnings fluctuate substantially from year to year, depending on system conditions and unexpected events. This can be especially true in an energy-only market. For example, the period from 2021 to 2023 reflected the latter, marked by elevated prices due to Winter Storm Uri and the inefficient procurement of the ERCOT Contingency Reserve Service (ECRS) in 2023. These years provided strong revenues for many generators, helping to offset earlier or future periods of lower earnings. However, 2024 exhibited lower net revenues as market conditions normalized and prices were more efficient.

Figure 56 presents historical net revenues available to support investment in new natural gas combustion turbines (CTs) and combined cycle (CC) generators.[48] These technologies are commonly considered the marginal new supply, meaning they are the types of units most likely to be built when the market signals a need for additional capacity. We calculate energy net revenues using generation-weighted real-time settlement point prices, assuming each unit sells energy or ancillary services in any hour it is economically profitable to do so.

Figure 56: Combustion Turbine (CT) and Combined Cycle (CC) Net Revenues, 2020-2024



48 For purposes of this analysis, we used the following assumptions: heat rates of 7 MMBtu per MWh for a combined-cycle unit, 10.5 MMBtu per MWh for a gas turbine, and $4 per MWh in variable operating and maintenance costs. A total outage rate (planned and forced) of 10% was assumed for each technology. It does not include: 1) start-up and minimum energy costs; or 2) ramping restrictions that can prevent generators from profiting during brief price spikes.

The CONE values used in this report are generated by Potomac Economics using the most current data publicly available. We use data specific to the ERCOT market and also leverage observations from other RTOs in the United States. The CONE calculation, by nature of the underlying formula, is very sensitive to certain inputs outside of the direct cost of installing a new generation plant. Values for weighted average cost of capital (WACC), the discount rate used to relate future cost into current dollar values, and the period over which the financial assessment is performed all have a pronounced impact on the calculated CONE value. There are notable differences among these values in the Potomac Economics model compared to the Brattle model which produced the CONE value that was approved by the PUCT in 2024.

1. Interpreting Single-Year Net Revenues

In 2024, marginal gas generators did not earn enough to cover their annualized capital costs. The CONE for CTs ranged from $102 per MWh to $106 per MWh, while their net revenues averaged only $68 per MWh. Similarly, CC units had a CONE between $116 per MWh and $121 per MWh, with average net revenues of $89 per MWh. While this shortfall may seem concerning in isolation, years like 2024 occur often and are not a threat to resource adequacy because one should expect tighter conditions to occur in other years that can produce revenues substantially above CONE.

If ERCOT's elevated load forecast discussed later in this chapter materializes, rising demand could again result in tighter conditions and stronger net revenue years. As always, single-year revenue results should be understood in the context of longer-term investment cycles. Generation developers are forward looking regarding revenue expectations and will often engage on forward contracting based on these expectations to lock in revenues and support financing for new projects. Hence, while historical net revenues may provide an empirical benchmark, investment will be driven by forward-looking expectations of load growth, generation development, interconnection costs, and market design changes that together determine profitability of new investment.

D. Profitability of Additional Resource Classes

In addition to discussing the profitability of natural gas in the context of CONE in the previous section, we also discuss nuclear energy and coal in this subsection. IRRs and ESRs have been treated at length in other chapters of this report.

Nuclear Energy. According to the Nuclear Energy Institute's "Nuclear Costs in Context" report, the average total generating cost for nuclear energy was $31.76 per MWh in 2023 and have been relatively stable.[49] The total generating cost is composed of: fuel costs, capital costs, and operating costs. Operating costs, which include expenses related to maintenance, security, and

[49] https://www.nei.org/resources/reports-briefs/nuclear-costs-in-context

labor, remained nearly flat at $19.38 per MWh, making up the largest portion of total costs. Plants with multiple units benefit from an economy of scale and outperform the average, particularly in operating costs. Given that the average zonal energy prices in 2024 ranged from $29.58 to $35.33 per MWh throughout the ERCOT market, typical nuclear resources would likely have covered their costs but may not have been profitable. As discussed above, however, prices and revenues can fluctuate substantially from year to year so it is not clear that nuclear resources would be unprofitable in the long term.

Coal. According to a December 2023 report from the U.S. Energy Information Administration (EIA), variable O&M costs for coal-fired power plants, excluding fuel, amount to approximately $6.40 per MWh. The same report estimates fixed operations and maintenance (O&M) costs at $61.60 per kW-yr,[50] which, assuming an 85% capacity factor, translates to roughly $8.28 per MWh. In ERCOT, the average fuel cost for coal-fired resources is roughly $8.50 per MWh.[51] Combined, this implies that the marginal production cost of coal-fired units in ERCOT is around $23.18 per MWh, which is lower than the national average largely due to the availability of low-cost Powder River Basin (PRB) coal. Given the prevailing energy prices in ERCOT in 2024, coal resources would have been profitable to run in many hours.

E. Peaker Net Margin

We reiterate that net revenue is the primary driver of resource adequacy, as it determines whether generators have sufficient incentive to invest or remain in the market. While this revenue often arises from shortage conditions, it can also result from broader supply and demand dynamics, such as the artificial shortage created by ECRS procurement in June 2023. Generators choose to enter the market when they anticipate that future market conditions will allow them to earn a sufficient return on investment. However, there are instances, like Winter Storm Uri in 2021, when single-year revenues spike well beyond what is necessary to support new entry should that revenue level be sustained. A resource owner is seeking an average net revenue across the investment horizon that will provide a positive return on the investment. This involves a mix of low and high net revenue years. If net revenue is too volatile over a period of years, even high net revenue years may not produce new investment due to perceived risk by the developer. In this situation, extremely high net revenue years may produce excess cost without an investment

---

50 https://www.eia.gov/analysis/studies/powerplants/capitalcost/pdf/capital_cost_AEO2025.pdf

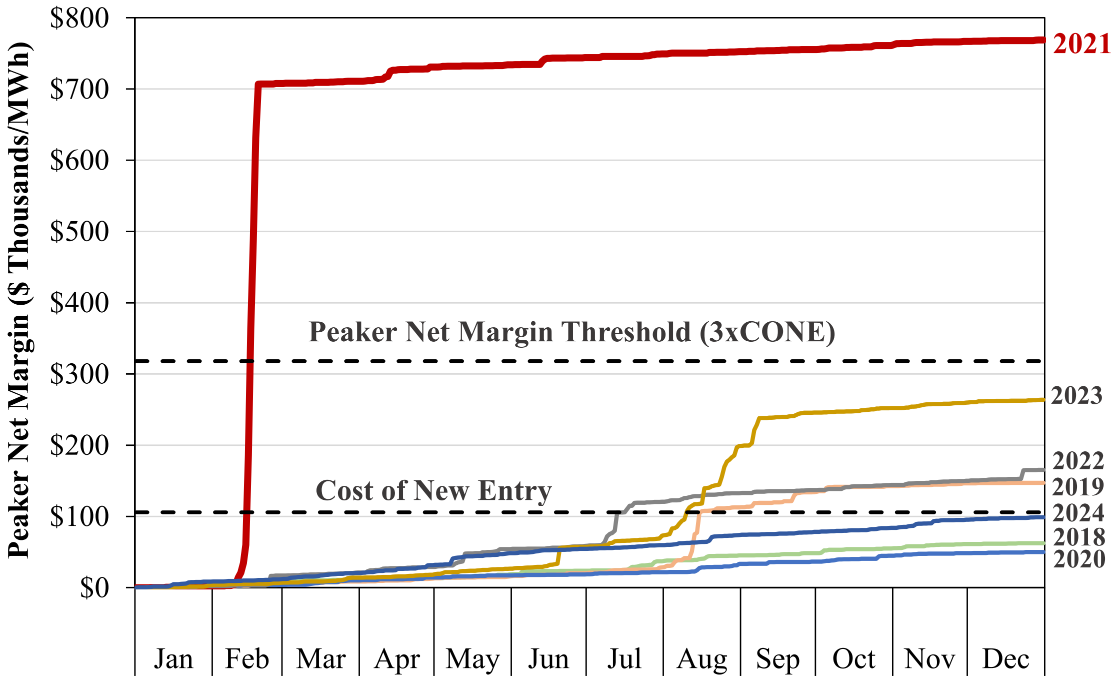
51 In ERCOT, most coal is imported from the Powder River Basin (PRB), where the average coal price in 2024 was about $0.80 per MMBtu. The average heat rate of coal-fired power plants has remained relatively stable over the past decade, increasing slightly from 10.4 MMBtu per MWh in 2013 to 10.7 MMBtu per MWh in 2023 as the national fleet continues to age. Based on a 10.7 MMBtu per MWh heat rate and $0.80 per MMBtu fuel cost. See https://www.eia.gov/electricity/annual/html/epa0801.html and https://www.eia.gov/coal/markets/#tabs-prices-2.

response to the price signal. To manage this dynamic, ERCOT uses the PNM threshold as a safeguard as outlined in TAC §25.509 of the PUCT's Electric Substantive Rules.[52]

1. Peaker Net Margin Threshold

PNM serves as a simplified benchmark for the annual net revenue that a hypothetical gas peaking unit could earn in the ERCOT market.[53] If, over the course of a calendar year, PNM exceeds a threshold of three times the CONE, equivalent to $315,000 per MW-year, the System-Wide Offer Cap (SWCAP) is reduced from $5,000 per MWh to $2,000 per MWh for the remainder of that year. This mechanism is designed to limit excessive shortage pricing once investment signals are deemed sufficient. Notably, this threshold has been exceeded only once in ERCOT's history, on February 16, 2021, during Winter Storm Uri. Figure 57 shows the PNM values for the past seven years.

Figure 57: Peaker Net Margin, 2018-2024



It is important to note that our net revenue calculation differs from ERCOT's PNM methodology. In our analysis, we assume a heat rate of 10.5 MMBtu per MWh, include variable O&M costs of $4 per MWh, and apply a total outage rate of 10%, which we believe reflects a more realistic estimate of generator performance and costs. By contrast, ERCOT's PNM calculation uses a simplified approach with a 10.0 MMBtu per MWh heat rate and excludes both

52 https://ftp.puc.texas.gov/public/puct-info/agency/rulesnlaws/subrules/electric/25.509/25.509.pdf

53 The proxy combustion turbine in the Peaker Net Margin calculation assumes a heat rate of 10 MMBtu per MWh and includes no other variable operating costs or startup costs.

variable O&M costs and outage rates. As a result, our calculation in Section C of this chapter produces values lower than those derived using ERCOT's PNM methodology.

F. ERCOT's Reliability Standard

1. Background

A reliability standard prescribes a level of supply that is required in order to meet certain reliability criteria. The assessment that leads to a prescribed standard, in terms of installed capacity, assesses potential reliability under various system conditions including more extreme conditions as can be experienced during winter cold snaps and summer heat waves. Generally, it serves as a benchmark for determining if there is sufficient capacity in the system for reliable operation. If a reliability standard is mandatory, it serves as imposed demand for installed capacity that can drive new investment in periods when the standard is not otherwise met.

The foundation for ERCOT's reliability standard was established through Senate Bill (SB) 3, passed by the 87th Texas Legislature in the aftermath of Winter Storm Uri in 2021. Among its wide-ranging reforms to improve electric grid resilience, SB 3 directed the PUCT to develop and implement a formal reliability standard for the ERCOT power region. This legislative mandate recognized the need for a clearer definition of acceptable system reliability and the mechanisms by which it should be evaluated. In response, the PUCT adopted 16 Texas Administrative Code (TAC) §25.508, which formalizes a probabilistic reliability standard based on loss of load expectation (LOLE), along with additional criteria for the duration and magnitude of load shed events.[54] In this section, we will introduce the requirements of the reliability standard and its importance to resource adequacy, summarize the analysis we submitted in July 2024, [55] and reiterate the proposal that arose out of this analysis.

2. TAC §25.508 Requirements

The new rule establishes a formal probabilistic reliability standard for the ERCOT system, structured around three key metrics:

- **Frequency:** The LOLE must be no greater than 0.1 events per year, or one event every ten years, on average.
- **Duration:** The maximum expected duration of a loss of load event must be less than 12 hours, with a 1.00% exceedance tolerance.

---

54 The PUCT organized the Reliability Standard under Project 54584, found here: https://interchange.puc.texas.gov/Search/Filings?ControlNumber=54584

55 Our comments are filed under: https://interchange.puc.texas.gov/search/documents/?controlNumber=54584&itemNumber=91

- **Magnitude:** The expected highest hourly level of load shed must be less than the amount of load that can be safely rotated, as determined by ERCOT, also with a 1.00% exceedance tolerance. ERCOT is required to annually determine and file the maximum amount of load that can be safely rotated during an event, along with the methodology used to calculate it.

Starting in 2026, ERCOT must conduct a reliability assessment at least once every three years to determine whether the system meets the standard and is likely to continue meeting it for the following three years. If the assessment shows the standard is not met, ERCOT must propose potential market design changes, and we (the IMM) must independently review those proposals.

3. Importance of a Reliability Standard

It is important to discuss the reliability standard because it directly shapes how ERCOT evaluates whether the system has sufficient resources to meet demand under a range of conditions. By formalizing a standard for loss of load events, along with limits on their duration and magnitude, the rule introduces clear expectations for system performance. This, in turn, has implications for resource adequacy, as the market must ensure that enough capacity and operational flexibility are available to meet the standard. Understanding how the reliability standard interacts with market design and investment signals is essential to evaluating whether ERCOT's resource mix can continue to deliver reliable service.

The reliability standard is closely tied to the concept of the VOLL, which represents the economic cost customers incur when electricity service is interrupted. In effect, the standard sets an implicit level of reliability that the system must deliver, and the cost of meeting that standard reflects the value society places on avoiding outages. If the standard is too stringent, it may imply a VOLL far higher than what customers are actually willing to pay, leading to excessive investment or market interventions. Conversely, a weak standard could result in underinvestment and unacceptable reliability outcomes. For this reason, it is important to ensure that the adopted reliability standard reflects a reasonable and economically justified VOLL, balancing the cost of reliability with the cost of outages.

4. Reliability in ERCOT

ERCOT operates as an energy-only market, meaning that new investment is incentivized primarily through shortage pricing. Higher prices during tight supply conditions signal the need for additional resources. However, this creates a fundamental tension: in order to trigger investment signals, the system must approach conditions of shortage, which inherently threatens reliability. To reconcile this, the VOLL must be set extremely high to support both effective shortage pricing and an ambitious reliability standard. This challenge is one reason many other electricity markets rely on capacity markets, which provide separate payments to ensure long-term resource adequacy.

In our 2022 State of the Market Report, we estimated that ERCOT's shortage pricing mechanism, based on the ORDC, implies a VOLL of approximately $47,000 per MWh. This is significantly lower than the roughly $200,000 per MWh implied by a typical 1-in-10 reliability standard or the presumably materially higher cost per MWh needed to support a reliability standard like the one adopted by the Commission.

5. Concerns and Conclusions

A single reliability standard based on Expected Unserved Energy (EUE) offers important advantages over a multi-factor standard. EUE captures both the duration and magnitude of potential loss-of-load events, meaning it inherently reflects the severity and length of outages in a single, unified metric. This makes it well-suited for aligning with a reasonable VOLL and simplifies both modeling and implementation. In contrast, applying three separate mandatory standards, for frequency, duration, and magnitude, can over-prescribe installed capacity and result in implied VOLLs far exceeding reasonable levels. In some cases, the implied VOLL from trying to meet a binding magnitude or duration constraint could exceed $1M per MWh, far beyond what most consumers would be willing to pay to avoid outages.

Our primary concern lies with the magnitude standard, which we believe is the most volatile and sensitive to planning model assumptions. Small changes in load forecasting methods, resource modeling, or historical weather scenarios can sharply alter the estimated peak load shed, potentially making the magnitude standard binding even when the overall system is adequate. We also find that the duration standard adds an unnecessary boundary condition without providing meaningful reliability value on its own. In our original comments to the Commission, we recommended eliminating both the magnitude and duration criteria and replacing them with a single EUE-based standard that would more effectively reflect reliability risk and better support economically justified planning.

While the Commission ultimately adopted the three-standard approach, it did take steps to address some of our concerns by relaxing the exceedance tolerances associated with the duration and magnitude metrics, which helps mitigate their potential to drive excessive investment. Since the current reliability standard may have significant impacts on generators and consumers alike, we advise clarifying formally whether the Commission intends the reliability standard to serve as a binding objective for ERCOT or remain an aspirational target to inform market participants. Absent this clarification, we interpret the directive to ERCOT to produce proposals to meet the standard as an indication that the standard is or will be mandatory.

G. Communicating Resource Adequacy

The way resource adequacy is communicated plays a critical role in shaping the expectations and decisions of market participants. Generators, LSEs, and other stakeholders rely on ERCOT's resource adequacy assessments to inform long-term investment strategies, operational planning, and risk management. While individual market participants may interpret market signals differently based on their own risk tolerances and business models, the formal reports and forecasts published by ERCOT serve as widely referenced benchmarks. These reports help frame expectations around future supply and demand conditions, influencing market behavior in tangible ways.

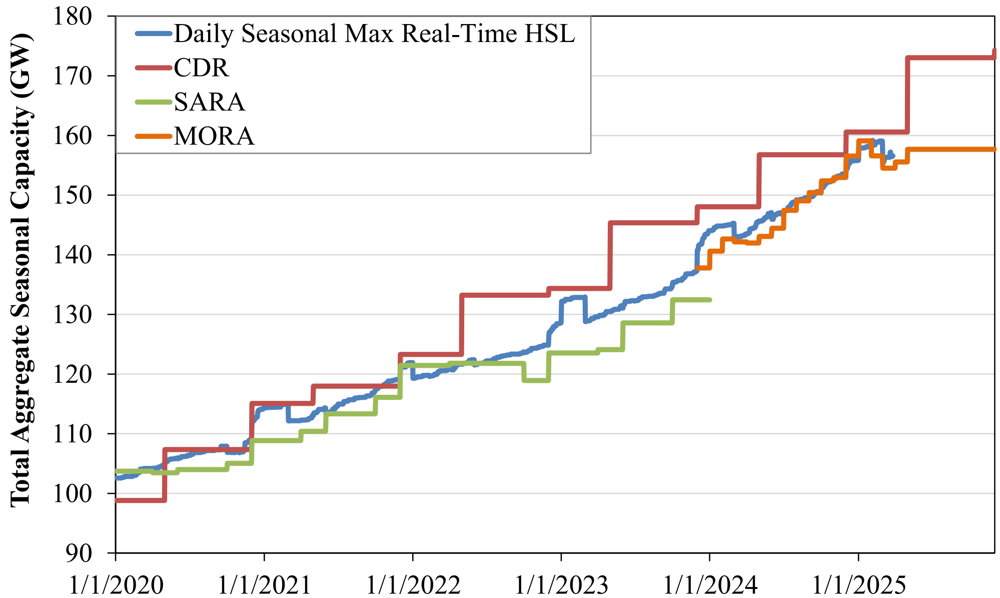
In this section, we review the primary reports through which ERCOT communicates resource adequacy to the market. We will discuss how ERCOT's Monthly Outlook for Resource Adequacy (MORA), the Long-Term Load Forecast (LTLF), and the Capacity, Demand, and Reserves (CDR) report convey expectations about system conditions over varying time horizons. We will also highlight key limitations in ERCOT's current resource adequacy communications.

1. Monthly Outlook for Resource Adequacy

The MORA report provides an early assessment of the risk that ERCOT may need to issue an Energy Emergency Alert (EEA) or initiate controlled outages during the reporting month. Introduced in 2023 to replace the Seasonal Assessment of Resource Adequacy (SARA), the MORA offers more granular, month-ahead insights. It uses probabilistic modeling to estimate the likelihood of insufficient operating reserves during peak demand periods and includes scenarios showing expected demand and resource availability. The MORA is particularly useful for forecasting resource adequacy conditions in the near term, typically one to two months in advance, and should be viewed as a short-term planning tool rather than a long-term resource adequacy assessment.

Figure 58 compares the short-term forecast accuracy of the MORA, SARA, and CDR reports against the actual Daily Seasonal Max Real-Time HSL, which represents the total online capacity approved to operate at maximum output on any given day. Among the three, the MORA shows the closest alignment with actual conditions, consistently tracking the Daily Max HSL throughout the year. By contrast, the SARA report, which was last published in 2023, showed less accuracy, in part because it was only updated on a seasonal basis and could not capture month-to-month changes in resource availability.

Figure 58: Short-Term Forecast Accuracy, MORA, SARA, & CDR



2. Capacity, Demand, and Reserves Report

The CDR report is ERCOT's primary long-term resource adequacy forecast. Published twice a year, the CDR provides a five-year outlook of expected system conditions by projecting future capacity additions and retirements alongside forecasted peak demand. The report is designed to inform market participants, regulators, and stakeholders about whether anticipated resources will be sufficient to meet projected demand under normal weather conditions. Unlike the MORA, which is a short-term operational tool, the CDR serves as a planning tool to provide a broad view of resource adequacy trends over the medium term. As a result of its focus on long-term forecasts, the CDR is less effective at communicating short-term resource adequacy, a limitation that is evident in its reduced accuracy compared to the MORA in Figure 58.

Over the past year, the CDR report has undergone significant improvements aimed at providing a more accurate and realistic assessment of future resource adequacy. NPRR 1219, implemented in October 2024, implemented a series of improvements to the CDR report.[56] One of the most notable changes is the shift from using peak average capacity contribution to the more sophisticated Effective Load Carrying Capability (ELCC) methodology for variable resources like wind and solar. The CDR also incorporates the ELCC of ESRs, where ELCC increases non-linearly based on system duration as larger systems provide diminishing returns. This change better reflects the contribution of these resources to reliability, particularly as their penetration has grown. The report now also includes planned retirements of generation resources in its forecasts, providing a clearer picture of potential supply shortfalls in future years.

[56] NPRR 1219 introduced several improvements to the CDR report and we only cover a couple of them here. The comprehensive language can be found here: https://www.ercot.com/mktrules/issues/NPRR1219

The shift from using peak average capacity contribution to ELCC in the CDR report is particularly significant because it accounts for the evolving timing of system stress. Specifically, it reflects the shift of the net peak load hour from the afternoon, when solar output is typically at its highest, to the evening hours, when solar generation diminishes or disappears entirely. This change highlights the growing importance of the net peak load, which measures demand after subtracting available renewable generation. Historically, the system's most strained conditions occurred at the firm peak load, the highest gross demand hour. However, as renewable penetration has increased, the hour of greatest reliability risk has shifted to the evening hours when solar output declines.

While the CDR report has undergone meaningful improvements in recent years, several significant deficiencies remain that limit its usefulness as a tool for stakeholders to assess long-term resource adequacy. These limitations can lead to an incomplete or misleading view of future supply and demand conditions, affecting how market participants plan for the future. The key deficiencies are as follows:

Excludes demand response resources: The CDR report does not account for the contribution of demand response. According to the November 2024 Annual Report on ERCOT Demand Response, there are approximately 10.5 GW of Non-Controllable Load Resources (NCLRs) and 1.2 GW of Controllable Load Resources (CLRs). Combined, these resources total 11.7 GW of flexible load that reliably reduces consumption during periods of system strain, the very conditions the CDR is intended to evaluate. During these events, these load resources will reliably reduce to a fraction of their peak demand.

Unreliable beyond two-year horizon: The CDR does not anticipate new generation resources that may enter the interconnection queue in the outer years of its five-year forecast horizon. Many of these resources could feasibly become operational by the end of the forecast period, yet they are excluded from the analysis. This limits the CDR's ability to accurately represent the evolving resource mix and leads to a conservative, and potentially misleading, view of long-term resource adequacy. We will discuss how generation is forecasted in an upcoming section.

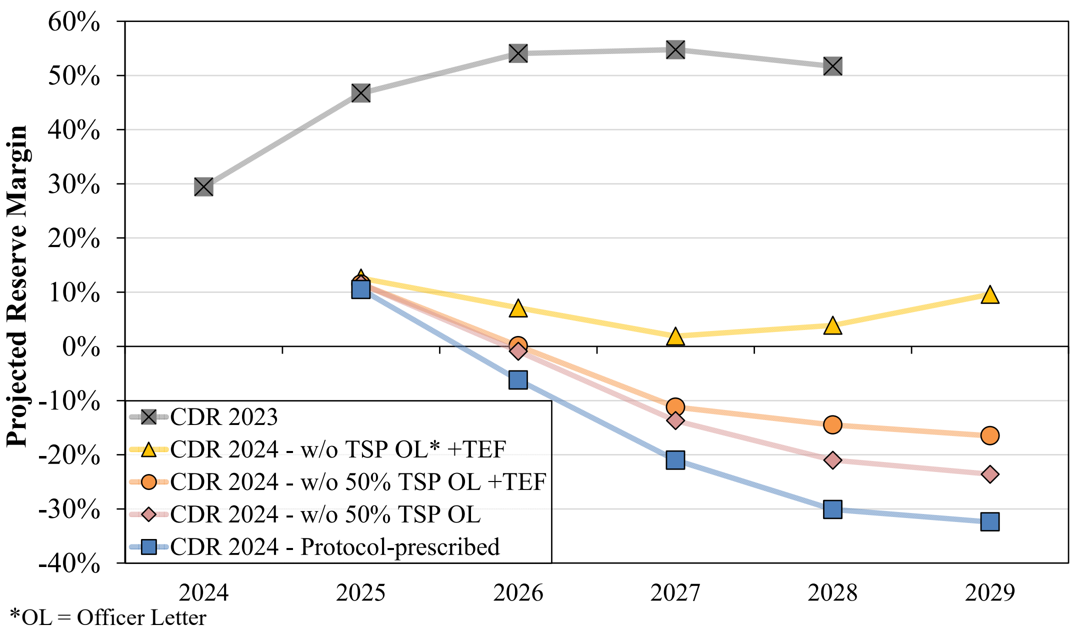
Relies on planning data: The CDR depends heavily on planning data submitted by market participants, including projected capacities and operational dates for future generation resources. This information is often incomplete or inaccurate, as it relies on voluntary self-reporting without reconciliation against actual operational outcomes. Without a systematic process to reconcile this planning data with operations data, the CDR's forecasts remain vulnerable to errors and may misrepresent the true state of resource adequacy.

3. Impact on Reserve Margin

The planning reserve margin is a metric used to measure the difference between the total available generation capacity and the expected peak demand, expressed as a percentage of peak demand. It serves as a simple indicator of whether there is sufficient capacity to meet demand under typical system conditions. A higher reserve margin indicates more available capacity, reducing the risk of shortages, while a lower margin suggests tighter supply conditions. A negative reserve margin means that projected generation capacity is insufficient to meet peak demand, increasing the likelihood of reliability concerns, price spikes, and potential load-shedding events.

Recent changes to the CDR report have had a substantial impact on projected reserve margins. While the December 2023 CDR projected reserve margins ranging from 27% to 51.7% five years out,[57] the December 2024 CDR projects a much lower range of -32.4% to 9.6% over the same horizon.[58] This dramatic decline is driven by two primary changes in the CDR methodology. First, the shift to using ELCC reduces the contribution of renewable resources during net peak load conditions, leading to lower counted capacity. Second, ERCOT now mandatorily includes TSP officer letter loads in its load forecast, significantly increasing the expected demand in future years. Figure 59 compares the reserve margin projections between the 2023 and 2024 CDR reports.[59]

Figure 59: Planning Reserve Margin, CDR 2023 vs CDR 2024



In the Load Forecast subsection, we examine in more detail what transmission service provider (TSP) officer letter loads are and how ERCOT's inclusion of them has affected the load forecast

57 https://www.ercot.com/files/docs/2023/12/07/CapacityDemandandReservesReport_Dec2023.pdf

58 https://www.ercot.com/files/docs/2025/02/12/CapacityDemandandReservesReport_December2024.pdf

59 ERCOT produced the May 2025 CDR just prior to the publication of this report. The reserve margins across all considered scenarios vary between -20.6% and -50.3% for the peak net load hour in the summer. It can be found here: https://www.ercot.com/files/docs/2025/05/15/CapacityDemandandReservesReport_May2025.pdf

and, by extension, the reserve margin. First, however, we will finish out this section with a brief discussion of the final report that ERCOT uses to communicate resource adequacy, the LTLF.

4. Long-Term Load Forecast

The LTLF is published yearly and serves as ERCOT's primary tool for projecting system load growth over an extended period, typically covering a ten-year horizon.[60] The LTLF provides detailed forecasts of future peak demand and energy consumption, using econometric models informed by economic and demographic trends, weather data, and other variables. The forecast includes six major components: base economic load, large flexible loads (LFLs), electric vehicle (EV) load, rooftop photovoltaic (PV) generation, contracted large industrial loads, and officer letter loads. Base load is forecasted using traditional economic indicators such as population growth, housing stock, and employment data, while other categories, including EVs and rooftop PV, are modeled based on adoption trends and historical patterns.

Despite its comprehensive scope, the LTLF report has limitations that affect its usefulness as a forecasting tool. Most notably, while contracted and TSP officer letter loads comprise the vast majority of the anticipated new load in ERCOT over the next several years, the LTLF does not evaluate the likelihood of these loads arriving within the forecast period. Instead, it includes these loads in full once they are reported, regardless of the development risks, permitting challenges, or other barriers to realization they face. ERCOT's 2025 LTLF adjustment to the loads submitted by TSPs is an improvement over its process in the 2024 LTLF but remains based on limited experience and data related to officer letter loads, making a data-driven approach challenging. This approach, combined with the narrow scope of information ERCOT receives from TSPs through its annual request for information (RFI) process and the lack of a more robust history of the proportion of stated load expectation actually becomes commercial, limits the potential accuracy of future resource adequacy. ERCOT did evaluate alternative scenarios in the most recent CDR report, which provided additional value to developers, consumers, and policy makers. Further exploration of how to better reflect uncertainty around the largest drivers of future demand will provide even more accurate and valuable projections.

H. Load Forecast

ERCOT's load forecasts have changed dramatically over the past two years, reflecting new policies, evolving methodologies, and a rapidly shifting load landscape. Historically, peak demand in ERCOT grew at a steady pace, increasing from 74.7 GW in 2019 to 85.6 GW in 2024, for an average annual growth rate of 2.76%.[61] However, recent forecasts have diverged sharply from this trend, with ERCOT projecting much higher levels of future demand. This section reviews the sequence of peak load forecasts ERCOT has published in the past two years,

60 https://www.ercot.com/files/docs/2024/01/18/2024_LTLF_Report.docx

61 https://www.ercot.com/gridinfo/load/load_hist

explains how policy changes such as the inclusion of officer letter loads have contributed to these shifts, identifies the persistent limitations in ERCOT's forecasting process, and considers the scale of transmission investment that may be required if these forecasts materialize. Table 9 summarizes the range of load forecasts ERCOT has published since December 2023, through both the CDR and LTLF reports.

Table 9: ERCOT's Load Forecasts Summary

	Firm Peak Load (GW)				Net Peak Load (GW)			
Report	**2024**	**2025**	**2029**	**2030**	**2024**	**2025**	**2029**	**2030**
Dec 2023 CDR	80.0	80.6	85.5	86.5	--	--	--	--
May 2024 CDR	--	80.6	82.7	82.8	--	--	--	--
July 2025 LTLF	86.0	90.5	140.9	148.0	--	--	--	--
Dec 2024 CDR	--	86.7	137.1	--	--	83.0	137.1	--
April 2025 LTLF TSP Provided	--	93.3	195.7	207.0	--	--	--	--
April 2025 LTLF ERCOT Adjusted	--	85.4	127.9	138.0	--	--	--	--
May 2025 CDR	--	--	128.3	138.4	--	--	115.9	126.6

1. Load Forecast Chronology

We begin by presenting a chronology that illustrates how ERCOT's forecasts have diverged over time as different data, methodological aspects, and observation have been incorporated over time.

June 2021. The Texas Legislature passed SB 1281, which took effect in September 2021. SB 1281 introduced Section 37.056(c-1) of the Utilities Code, which required load forecasts to consider (1) historical load, (2) forecasted load, and (3) additional load seeking interconnection.[62]

December 2022. The PUCT, who had organized the implementation of SB 1281 under Project 53403 earlier that year, implemented 16 TAC §25.101(b)(3)(A)(ii)(II) which requires that loads in the load forecast be substantiated by quantifiable evidence of projected load growth.[63] In the order adopting the amendments to 16 TAC §25.101, the PUCT stated that it will give great weight to written documentation by a TSP to ERCOT that a given transmission line is needed to interconnect certain loads.

62 https://capitol.texas.gov/tlodocs/87R/billtext/html/SB01281F.htm

63 https://interchange.puc.texas.gov/search/documents/?controlNumber=53403&itemNumber=86

May 2023. NPRR 1180[64] and PGRR 107[65] are introduced into the stakeholder process. Together, they introduce and define “Substantiated Load” and “Unsubstantiated Load” in ERCOT’s protocols. Both NPRR 1180 PGRR 107 are approved in January 2025, and define Substantiated Loads to refer to loads submitted by TDSPs for planning purposes substantiated by (1) an executed interconnection or other agreement, (2) an independent third-party load forecast deemed credible by ERCOT, or (3) a letter from a TDSP officer attesting to such load, which may include load for which a TDSP has yet to sign an interconnection agreement. The latter of these criteria introduces the concept of officer letter loads, discussed in greater detail later in this section.

June 2023. The Texas Legislature passed House Bill 5066, which clarified the definition of “additional load seeking interconnection” in Section 37.056(c-1) to include load that has not yet executed an interconnection agreement.[66] HB 5066 took immediate effect.

November 2023. ERCOT initiated its RFI process, soliciting submissions from TSPs for Substantiated Loads to be considered in ERCOT’s long-term transmission planning and resource adequacy forecasts.

December 2023. ERCOT published its December 2023 CDR. This report does not include officer letter loads as the RFI process was not yet complete.

May 2024. ERCOT published its May 2024 CDR, which also does not include officer letter loads.[67]

July 2024. ERCOT published its revised 2024 LTLF report, whose forecast includes officer letter loads. It projects a peak demand of 148 GW in 2030, up from 86.5 GW and 82.8 GW projected by the December 2023 CDR and May 2024 CDR reports, respectively.

November 2024. ERCOT issued its annual RFI to TSPs.

January 2025. NPRR 1180 and PGRR 107 are approved by the PUCT. The most significant distinction between the originally introduced version and the approved version is the clarification made by HB 5066, discussed earlier.

February 2025. ERCOT published its revised December 2024 CDR report. This report includes officer letter loads, including scenarios where some of the projected load may not materialize.

---

64 https://www.ercot.com/mktrules/issues/NPRR1180

65 https://www.ercot.com/mktrules/issues/PGRR107

66 https://capitol.texas.gov/tlodocs/88R/billtext/html/HB05066H.htm

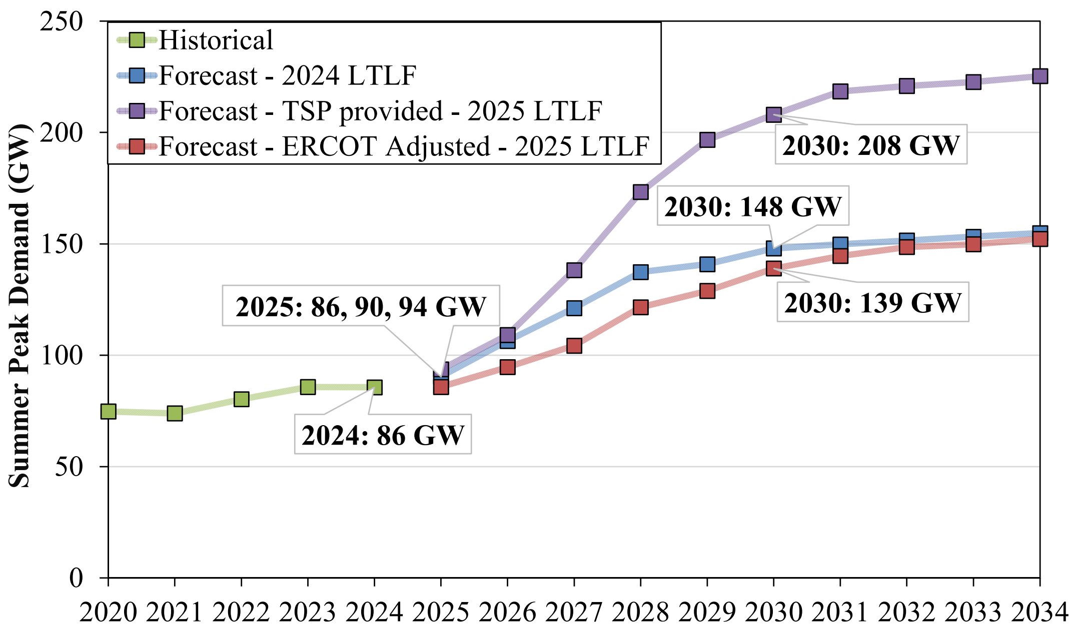
67 https://www.ercot.com/files/docs/2024/05/24/CapacityDemandandReservesReport_May2024_Revised.pdf

April 2025. ERCOT published its 2025 LTLF report. The TSP-provided load forecast for 2030 balloons to a summer peak demand of 208 GW. ERCOT also publishes an adjusted forecast that discounts the projected load to a summer peak demand of 138 GW.

May 2025. ERCOT filed with the PUCT a request for a good cause exception that would allow ERCOT to use its adjusted forecast for planning purposes in its 2025 regional transmission plan (RTP). As of this report's publication, this request has not yet been approved and is pending. ERCOT also published its May 2025 CDR.

To sum up, the 87th and 88th legislative sessions introduced new load forecasting requirements that have significantly inflated ERCOT's peak demand forecasts for the near future. These forecasts suggest that ERCOT will need to accelerate investments in new transmission projects and upgrades to keep pace with projected demand. The 2024 LTLF, 2025 ERCOT Adjusted Forecast, and 2025 TSP Provided Forecast show load growth rates of 9.7%, 8.5%, and 16.2%, respectively. These growth rates are three to six times higher than the historical growth rate of 2.76%. Figure 60 illustrates the peak demand forecasts as published in ERCOT's 2024 and 2025 LTLF reports.

Figure 60: ERCOT Long-Term Peak Demand Forecast



Next, we review how the net load forecast is calculated.

2. Net Forecast Calculation

The LTLF employs the following equation to calculate its net load forecast:

$$Net\ Forecast = Base\ Load\ Forecast + EV\ Forecast + LFL\ Forecast - PV\ Forecast + Contracted\ Load + Officer\ Letter\ Loads$$

ERCOT's long-term load forecast builds a system-wide net forecast by combining six categories: the base economic forecast, electric vehicle (EV) forecast, rooftop photovoltaic (PV) forecast, large flexible load (LFL) forecast, contracted load forecast, and officer letter load forecast. These components are layered using a waterfall approach. The base economic forecast is developed from econometric models, while EV forecasts use registration data mapped to ZIP codes. The rooftop PV forecast estimates peak demand reductions by customer class and weather zone. LFL projections account for typical curtailment during peak periods, and weather variability is considered for normal and probabilistic scenarios. The 2025 LTLF projects 2,006 MW of peak demand from EVs and a 2,083 MW peak demand reduction from rooftop PV by 2030.

The 2025 LTLF marks the first time ERCOT applied adjustment factors to contracted and officer letter load forecasts, based on observed delays and realization rates. This is an improvement over past practices, where such loads were included in full without adjustments.

3. Contracted Loads vs Officer Letter Loads

Contracted loads refer to large loads that have entered into interconnection agreements with a transmission service provider, making them relatively certain to come online. In contrast, officer letter loads are based on letters submitted by TSP officers expressing the intent to interconnect, but they have not yet finalized contractual commitments. As a result, officer letter loads are considered less certain and may carry a higher risk of delay or cancellation. Table 10 summarizes the contracted and officer letter loads used in ERCOT's 2024 and 2025 LTLF reports.

Table 10: 2024 & 2025 RFI Data, 2030 Forecast (MW)

LTLF By 2030	Type	Crypto/ LFL	Data Center	Hydrogen/ Ammonia	Oil & Gas	Industrial	Total
TSP Provided - 2024	Contracted	3,543	10,301	3,100	650	1,119	18,713
	Officer Letter	2,335	17,363	13,945	1,042	2,214	36,898
	Total	5,878	27,664	17,045	1,692	3,332	55,611
TSP Provided - 2025	Contracted	4,920	11,885	4,100	2,623	3,179	26,707
	Officer Letter	6,402	66,081	8,862	917	7,885	90,146
	Total	11,321	77,965	12,962	3,540	11,064	116,853
ERCOT Adjusted - 2025	Contracted	4,176	5,746	4,100	2,617	2,921	19,560
	Officer Letter	3,325	16,429	3,802	508	4,159	28,223
	Total	7,500	22,175	7,902	3,125	7,080	47,783

We highlight a few key statistics derived from Table 10:

- The forecasted load submitted by TSPs more than doubled from the 2024 RFI to the 2025 RFI. Nearly 80% of this increase, approximately 49 GW, came from growth in data center officer letter loads.
- Forecasted load in the Crypto/LFL and oil and gas categories roughly doubled between the 2024 and 2025 RFIs, while industrial load more than tripled.
- Planned Hydrogen/Ammonia projects saw significant delays and cancellations between the 2024 and 2025 RFIs, resulting in a net decrease of 4 GW across both categories.
- Officer letter loads make up 66.4% of forecasted load growth in the 2024 LTLF, 77.1% in the 2025 LTLF TSP Provided forecast, and 59.1% in the 2025 LTLF ERCOT Adjusted forecast.

4. Transmission Investment

ERCOT has historically invested an average of $3 billion per year in building or upgrading transmission infrastructure, with that figure rising to $3.78 billion in 2024.[68] If we assume that transmission investment scales proportionally with load growth, then meeting the projected 8.5%-16.2% year-over-year increase in peak demand over the next five years would require ERCOT to multiply its annual investment in transmission projects three to six times over. However, this is a simple linear extrapolation that does not account for the secondary effects of such rapid growth, particularly the increased demand for labor, materials, and equipment, all of which could strain supply chains and escalate costs even further.

It is also important to note that the cost of new transmission infrastructure in ERCOT is socialized across all market participants, in line with Texas' open access laws. Under this structure, entities seeking to interconnect are not directly responsible for the full cost of the transmission upgrades required to serve their load, which lowers the barrier to entry and supports Texas' reputation as an attractive market for investment. However, this model also means that the financial burden falls on existing ratepayers, who ultimately fund transmission expansion through regulated transmission charges.

5. ERCOT's April 2025 Load Forecast Adjustment

When ERCOT first incorporated officer letter loads into its 2024 LTLF, it did so without explaining what officer letters are or acknowledging that these loads are far less likely to materialize than contracted loads or other forms of load growth. By including all 37 GW of officer letter loads without any adjustment or explanation, the forecast surged to 148 GW,

68 https://www.ercot.com/files/docs/2025/01/27/2024-regional-transmission-plan-rtp-345-kv-plan-and-texas-765-kv-strategic-transmission-expans.pdf

implying an average annual growth rate of 9.7%. This sudden increase and lack of understanding of the data and methodology that produced it left stakeholders uncertain whether to treat the forecast as a credible signal for investment in new transmission or to discount it entirely.

In the 2025 LTLF, ERCOT introduced adjustment factors for the first time, using historical data to estimate how often officer letter loads actually materialize and energize. These adjustments delayed in-service dates for both contracted and officer letter loads by 180 days. They also reduced new data center demand to 49.8% of the originally requested capacity, and then further reduced all officer letter loads to 55.4% of the originally requested capacity. Contracted loads were not discounted besides the delay factor. These changes lowered the 2030 TSP Provided forecast from 208 GW to 139 GW.[69]

Understanding these adjustments requires distinguishing between two concepts: consumption and energization. Consumption refers to the peak demand a load actually uses compared to the MW it originally requested. Energization refers to whether a load comes online at all to begin drawing power.

ERCOT's adjustments were based on the following findings. As of January 31, 2025, 55.4% of officer letter loads had energized. Among the loads that did energize, only 22% of the originally requested capacity was actually being consumed. ERCOT discounted officer letter loads to 55.4% instead of 22% because it assumes that these loads may increase their consumption over time as operations ramp up. Effectively, ERCOT applied the energization rate to the consumption rate. For data centers, which account for the vast majority of forecasted load growth, ERCOT found that these facilities consumed only 49.8% of their requested capacity. This figure was based on a study of several large data centers over the past few years.

6. Validating the Forecast

It is difficult to validate ERCOT's forecast, and we do not attempt to do so here. Doing so would require access to detailed cost and transmission planning data that neither we nor ERCOT currently possess. The information necessary to assess the likelihood that a given load project will interconnect by a specific date resides with the individual TSPs. While a limited subset of this information is communicated to ERCOT through its annual RFI, the data is often incomplete and may quickly become outdated as project timelines shift. It is also important to understand that ERCOT has just over a year of data and experience with officer letter loads, which limits its ability to establish a fully data-driven approach and publish defensible discount factors.

Indeed, the communication chain from market participants to the public leaves room for gaps and distortions. In the interconnection process, a prospective load submits information to the

69 https://www.ercot.com/files/docs/2025/04/07/8.1-Long-Term-Load-Forecast-Update-2025-2031-and-Methodology-Changes.pdf

TSP, which then conducts its own internal assessment of what is needed to support interconnection. ERCOT's yearly RFI does not capture this detailed evaluation. Instead, it collects high-level information, such as whether an interconnection agreement has been signed, the location of the load, and whether it qualifies as an officer letter load, along with a few other general data points. These inputs do not provide sufficient basis to assess the likelihood or timeline of interconnection, limiting the transparency and reliability of the load forecast.

I. Generation Forecast

We conclude this chapter with a brief review of how generation capacity is forecasted and communicated in ERCOT. As previously discussed, generation forecasts are communicated exclusively through the CDR report, which serves as the primary source of forward-looking information on system capacity. The CDR calculates its generation forecast by combining current operational capacity with two additional categories: resources with signed interconnection agreements and those that are synchronized to the grid but not yet approved for commercial operation. This methodology allows the CDR to reasonably estimate the amount of capacity that is likely to be online within the next two years.

However, the CDR does not account for generation projects that have not yet signed an interconnection agreement, even if those projects are likely to come online in years three through five of the forecast horizon. As a result, the CDR's longer-term outlook tends to understate future capacity, especially beyond the two-year mark. Given these limitations, we emphasize that the CDR functions best as a medium-term forecasting tool, not a long-range planning document. Its attempt to forecast generation five years into the future can be misleading and unnecessarily alarming to stakeholders. This concern is particularly evident in the forecasted reserve margins shown in Figure 59, which reflect the limited scope of the generation forecast rather than a full picture of future system capability.

J. Conclusion

The discussion around resource adequacy is a critical one to get right, as it shapes both investment decisions by market participants and regulatory responses from the Legislature and the Commission. A mischaracterization of future system needs can lead to overinvestment, underinvestment, or misguided policy interventions. Below, we reiterate five key takeaways from this chapter.

First, the LTLF has functioned more as an accounting ledger than a forecasting model. It aggregates large volumes of anticipated load, primarily from contracted and officer letter loads, without applying adequate filters to assess the likelihood that these loads will materialize within the forecast period. While the current methodology captures what has been reported to ERCOT, it does not evaluate economic viability, permitting status, or development timelines. As a result, the forecast risks overstating future peak demand, which can distort perceptions of system adequacy and infrastructure needs. ERCOT's recent adjustment to its forecast improves this paradigm but is based on a limited data set that does not yet capture officer letter load behavior.

Second, the CDR suffers from similar limitations, particularly beyond its two-year horizon. While it reasonably captures generation and load additions already in motion, it fails to account for new resources that have not yet signed interconnection agreements but could feasibly enter commercial operations within the five-year forecast period. This structural limitation makes the CDR less a forward-looking tool and more a reflection of the current queue status. Its five-year reserve margin projections, as shown in Figure 59, can present a misleading picture of risk and may overstate the need for intervention.

Third, the weakness of the communication chain that underlies ERCOT's forecasting process. Load and generation data flows from market participants to TSPs, then to ERCOT through an annual RFI, and finally to the public through official forecasts. This process introduces multiple opportunities for information loss or misinterpretation. For example, ERCOT receives high-level data but not the detailed project status information held by individual TSPs, such as permitting progress or construction challenges, that would be necessary to assess the timing and viability of interconnection. The result is a forecast that lacks transparency and completeness.

The fourth takeaway concerns the uncertainty surrounding the reliability standard adopted in 2024. While the standard sets clear probabilistic benchmarks for system reliability, it remains unclear whether the Commission intends to treat it as a binding requirement or an aspirational guideline. Without that clarity, market participants may misinterpret the urgency or consequences of failing to meet the standard, which could lead to either overbuilding or delays in needed investment. This ambiguity could complicate planning and market design decisions moving forward.

Finally, if ERCOT's revised load forecast proves accurate, a substantial increase in transmission investment will be required to maintain reliability. ERCOT has historically invested around $3 billion per year in transmission upgrades, rising to $3.78 billion in 2024. But a 9.7%-16.2% annual increase in peak demand, as projected over the next five years, would require transmission investment to multiply by a factor of three to six. Since Texas' open access structure socializes transmission costs across all market participants, this shift could place a significant financial burden on existing ratepayers. A clear understanding of both the forecast and its implications is essential to preparing for this level of system expansion.

VII. Analysis of Competitive Performance

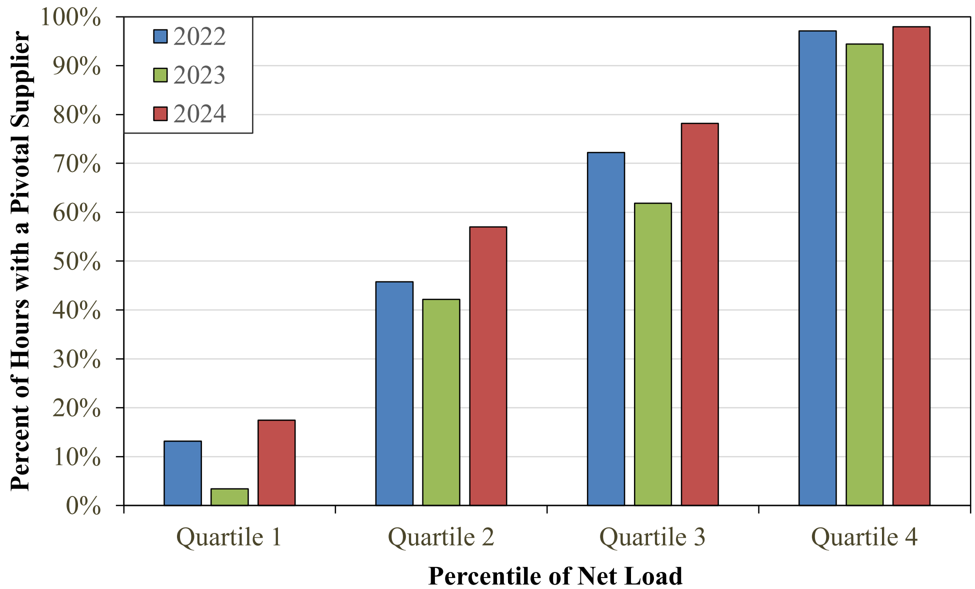
In this section, we evaluate market power from two perspectives: structural (does market power exist) and behavioral (have attempts been made to exercise it). This section begins by evaluating a structural indicator of potential market power, then evaluates market participant conduct by reviewing measures of potential physical and economic withholding. Finally, this section includes a high-level summary of the Voluntary Mitigation Plans (VMPs) in effect during 2024.

Based on these analyses, we find that the ERCOT wholesale markets generally performed competitively in 2024. However, our assessment of market power possessed by large suppliers and the extent of offer prices well in excess of competitive levels raises concern that may warrant additional scrutiny of VMP provisions and additional market power mitigation measures to capture instances outside of those facilitated by binding uncompetitive transmission constraints.

A. Structural Market Power Indicators

The market is most competitive when no participant can withhold the capacity, either physically or economically via high-priced offers, in order to benefit by raising prices substantially. Traditional market concentration measures are not reliable market power indicators in electricity markets partly because they do not consider load obligations that affect suppliers' incentives to raise prices. They also do not account for excess supply, which affects the competitiveness of the market. A more reliable indicator of market power is whether a supplier is "pivotal," i.e., whether a suppliers' resources are required to meet demand for energy and ancillary services or manage congestion. Figure 61 shows the results of our pivotal supplier analysis by showing the frequency of hours where there is a pivotal supplier, grouped by net load level (quartile).

Figure 61: Pivotal Supplier Frequency by Net Load Level



During the top 25th percentile of net load occurring in each year, there was a pivotal supplier greater than 90% of the time. Inherently, high net load indicates a large demand to be satisfied by generation resources (and net imports) not including wind and solar. System conditions, including winter cold snaps and summer heat waves, along with a high degree of generation outages can have a significant impact on the balance of supply and demand which can directly impact the extent to which participants possess market power. Decision Making Entities (DMEs) with a high portfolio of non-intermittent generation are an important contributor during these hours. The frequency of hours where there are pivotal suppliers is expected to increase with the level of net load. As load reserve requirements increase, there is less excess supply to meet those needs. This increases the potential market power of participants. Pivotal suppliers, market participants with structural market power, existed in 63% of all hours in 2024, compared to 50% in 2023 and 57% in 2022.

We also evaluate competitiveness at a zonal level. The methodology follows the same structure as the system-wide evaluation with two exceptions. First, the zonal approach does not consider reserve requirements at the zonal level. ERCOT does not have zonal demand curves for reserves, so there is no explicit requirement that some reserves be sourced within a specific zone. Second, import capability into a zone is competing with zonal supply and is addressed through netting observed net imports into a zone from the load in that zone. The figures in Table 11Table 11 show, by zone, the percentages of hours during the top quartile of load where one or more pivotal suppliers existed.

Table 11: Frequency of One or More Pivotal Suppliers in Top Quartile of Net Load by Zone

	Pivotal Frequency		
	2022	**2023**	**2024**
Houston	35%	52%	28%
North	65%	68%	66%
South	1%	0%	3%
West	8%	21%	50%

This analysis focuses on hours where the load level was in the highest 25% of the year, when supply is most likely to be tighter and lead to the potential exercise of market power. The Houston and North zones have the highest prevalence of structural market power in these high load hours. It is notable that the West zone also experienced a high frequency of structural market power in 2024. For perspective, there are 2,190 hours represented in the highest quartile of load throughout a year. That is 1,445 hours for the North zone in 2024, showing 66% of time with one or more pivotal suppliers. This high frequency of uncompetitive supply conditions provides considerable opportunity for a pivotal supplier to profitably increase price.

We cannot make inferences regarding market power solely from pivotal supplier data because it does not consider the contractual position of the supplier. Bilateral and other financial contract obligations can affect whether a supplier has the incentive to raise prices. For example, a small supplier selling energy solely in the real-time energy market may have a much greater incentive to exercise market power than a large supplier with substantial long-term sales contracts. We recommend that the "small fish" rule be eliminated because these small suppliers are sometimes pivotal, and because high offer prices are not necessary to ensure efficient pricing under tight conditions (see SOM Recommendation 2021-1).

It should be noted that the analysis above evaluates the structure of the entire ERCOT market. In general, local market power in smaller geographic areas of the power region that can become isolated by transmission constraints raise more substantial competitiveness concerns. As more fully discussed in Chapter IV, this local market power is addressed through: (a) structural tests that determine "non-competitive" constraints that can create local market power; and (b) the "mitigation" or application of limits on offer prices in these areas.

B. Evaluation of Supplier Conduct

This subsection provides the results of our evaluation of actual conduct to assess whether market participants have attempted to exercise market power through physical or economic withholding. First, we examine unit deratings and forced outages to detect physical withholding, and then we review the "output gap" used to detect economic withholding.

In a single-price auction like the real-time energy market, suppliers may attempt to exercise market power by withholding resources. The purpose of withholding is to cause more expensive resources to set higher prices, allowing the supplier to profit from its other sales in the market. Because forward prices are highly correlated with spot prices, price increases in the real-time energy market can also increase a supplier's profits in the bilateral energy market. This strategy is profitable if the incremental profit exceeds the foregone profits from its withheld capacity.

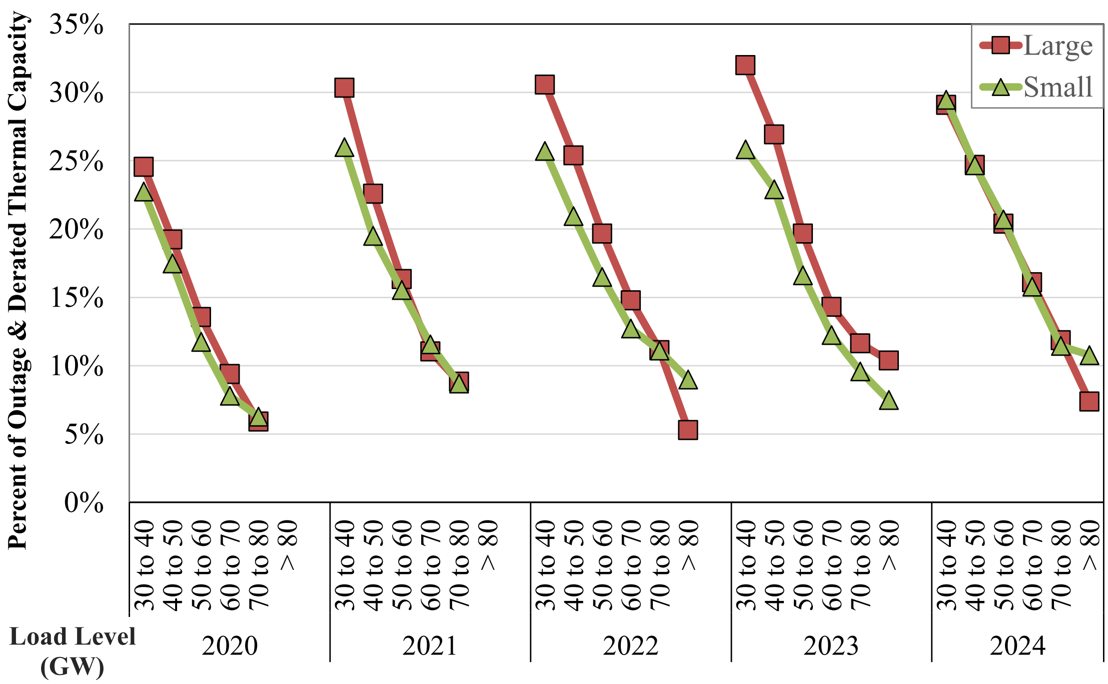
1. Evaluation of Potential Physical Withholding

Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and are economic at market clearing prices. A plant operator can withhold either by derating a unit or declaring the unit as forced out of service. Because generator deratings and forced outages are unavoidable, the goal of the analysis in this subsection is to differentiate justifiable deratings and outages from physical withholding. We conduct a test for physical withholding by examining deratings and outage data to ascertain whether the data are correlated with conditions under which physical withholding would likely be most profitable.

The pivotal supplier results shown in Figure 61 indicates that the potential for market power abuse rises at higher net load levels, as the frequency of intervals in which suppliers are pivotal increases. Hence, if physical withholding is occurring, one would expect to see increased deratings and outages at the highest load levels. Conversely, because competitive prices increase as load increases, deratings and outages in a market performing competitively will tend to decrease as load approaches peak levels. Suppliers that lack market power will take actions to maximize the availability of their resources because their output is generally most profitable in peak periods.

Figure 62 shows the average aggregate planned, forced, and unreported outages as a percentage of total installed capacity for large and small suppliers under different real-time load levels. Portfolio size is important in determining whether suppliers have incentives to withhold available resources. Hence, we compare the patterns of outages and deratings of large and small suppliers. It is important to consider the aggregate number of outages due to the high frequency of pivotal supplier hours.

Figure 62: Outages and Deratings by Load Level and Participant Size



Wind, solar, and energy storage resources (ESRs) also are excluded from this analysis because of the high variation in the availability of these classes of resources. The large supplier category includes the four largest suppliers (DMEs) in ERCOT. The small supplier category includes the remaining suppliers.

Figure 62 confirms the pattern we have seen since 2018 that as demand for electricity increases, all market participants generally make slightly more capacity available to the market by scheduling planned outages during low load periods. The fact that available capacity tends to be higher under the highest load conditions is particularly notable because rising ambient temperatures generally cause thermal units' capability to fall.

Because small participants generally have less ability to physically withhold capacity to profitably exercise market power, the outage rates for small suppliers serve as a good benchmark for competitive behavior expected from the larger suppliers. Outage rates for large suppliers at all load levels modestly exceeded those for small suppliers but remained at levels that are small enough to raise no competitiveness concerns.

2. Evaluation of Potential Economic Withholding

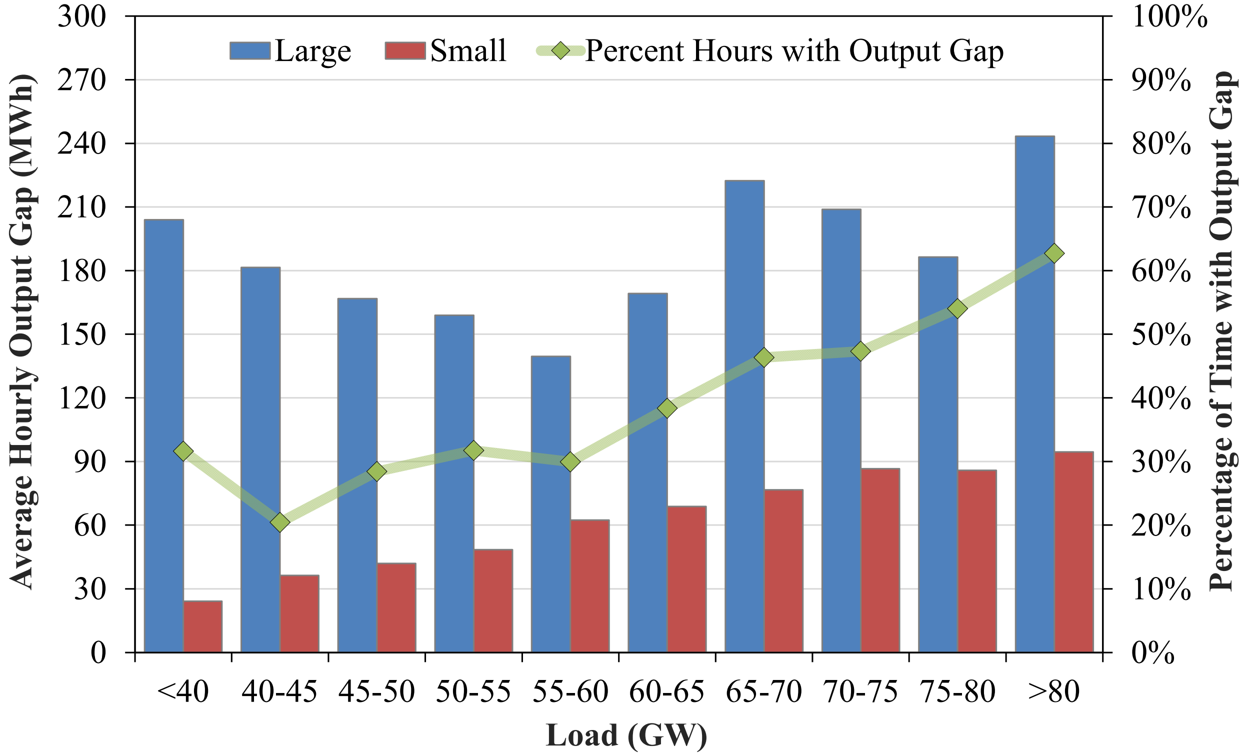
In this subsection, we evaluate potential economic withholding by calculating an "output gap". The output gap is the quantity of energy that is not being produced by online resources even though the output is economic to produce by a substantial margin given the real-time energy price. A participant can economically withhold resources, as measured by the output gap, by raising its energy offers for a resource to reduce its dispatch level.

Resources included in the output gap are those that are committed and producing at less than full output. Energy not produced from a committed resource is included in the output gap if the real-time energy price exceeds that unit's mitigated offer cap by at least $30 per MWh. The mitigated offer cap serves as a proxy for the marginal production cost of energy from that resource.

Figure 63 shows the average output gap levels, measured by the difference between a unit's operating level and the output level based on delivery over an hour had the unit been offered to the market based on a proxy for a competitive offer (i.e., the unit's mitigated offers), but with a few changes. We use generic costs instead of verifiable costs for quick-start units since verifiable costs may contain startup costs that are inappropriate for comparison here. In addition, fuel adders are removed since they represent fixed costs. Finally, we do include quick-start units if they were in quick-start mode and available for real-time dispatch. The information in Figure 63 reflects the average positive output gap by load level with the percentage of hours reflected in each load level category for reference.

In 2024, roughly 32% of the hours exhibited an output gap, indicating potential attempt to exercise market power through economic withholding. At higher load levels, an extremely small percentage of generating capacity exhibited an output gap for a large percentage of time. An even smaller percentage of generating capacity exhibited an output gap at lower load levels. These results show that potential economic withholding in the real-time energy market as low, but not trivial, in 2024. While the ERCOT market may have performed competitively in general, the level of market power and moderate evidence of potential economic withholding are cause for concern. Anticipated increase in system load over the coming years can result in more frequent structural market power and more incentive to exercise that market power.

Figure 63: Incremental Output Gap by Load Level and Participant Size



C. Voluntary Mitigation Plans

The PUCT has discretion to approve VMPs filed by market participants.[70] Before September 1, 2023, a market participant's adherence to a PUCT-approved VMP constituted an absolute defense against an allegation of market power abuse with respect to behaviors addressed by the plan. However, House Bill 1500, which was passed during the 88th Legislative session and went in effect on September 1, 2023, modified the statutory requirements related to VMPs. Adherence to a VMP is no longer considered an absolute defense against allegations of market power abuse with respect to the behaviors addressed by the VMP; instead, adherence to a VMP must be considered in determining whether a violation occurred and, if so, the penalty to be assessed.[71]

Generation owners are often motivated to enter into VMPs, and the increased regulatory certainty afforded to a generation owner regarding its energy offers in the ERCOT real-time market must be balanced by appropriate protections against a potential abuse of market power in violation of Public Utility Regulatory Act (PURA) §39.157(a) and 16 TAC §25.503(g)(7). In

70 PURA § 15.023(f).

71 *Id.* Also, the PUCT amended its rules to implement these statutory changes on April 25, 2024. *Review of Voluntary Mitigation Plan Requirements*, Docket No. 55948, Order (Apr. 25, 2024)

2023, Calpine, NRG, and Luminant had active and approved VMPs filed with the PUCT.[72] The PUCT modified these three VMPs on March 23, 2023 to address competitiveness concerns that the IMM raised in 2022 related to ERCOT's greatly increased procurement of Non-Spin Reserve Service (NSRS).[73] In February of 2024, NRG filed a letter with the PUCT expressing NRG's intent to exercise its right to terminate its VMP, effective March 1, 2024.[74]

The VMPs for Calpine and Luminant include provisions that specify competitive benchmarks for offers in both energy and reserves. Further, the provisions address different generation technologies and fuel types and also address on-line versus off-line states in consideration of competitive cost on which to base the offer cap. The IMM reviews the VMPs on a cycle and when significant changes to market rules may change the competitiveness of the market or one or more participants' degree of market power. Assessment and recommendations regarding VMP provisions are provided to PUCT staff.

VMPs should promote competitive outcomes and prevent abuse of market power through economic withholding in the ERCOT real-time energy market. The same restrictions are not required in forward energy markets (e.g., the ERCOT day-ahead market), but the prices in forward energy markets are informed by expectations for real-time energy prices (where mitigation is applied). The forward energy market is voluntary, and the market rules do not inhibit arbitrage between the forward energy market and the real-time energy market. Therefore, competitive outcomes in the real-time energy market serve to discipline the potential abuse of market power in the forward energy markets.

PURA defines market power abuses as "practices by persons possessing market power that are unreasonably discriminatory or tend to unreasonably restrict, impair, or reduce the level of competition."[75] The *exercise* of market power may not rise to the level of an *abuse* of market power if the actions in question do not unreasonably impair competition. Impairment of

---

72 *See Petition of Calpine Corporation for Approval of Voluntary Mitigation Plan,* Docket No. 40545, Order (Mar. 28, 2013); *Request for Approval of a Voluntary Mitigation Plan for NRG Companies Pursuant to PURA § 15 023(f) and P.U C. Subst. R. 25.504(e),* Docket No. 40488, Order (Jul. 13, 2012); *Request for Approval of an Amended Voluntary Mitigation Plan for NRG Companies,* Docket No. 42611, Order (Jul. 11, 2014); *PUCT Staff Request for Approval of a Voluntary Mitigation Plan for Luminant Energy Company, LLC under PURA §15.023(f) and 16 TAC §25.504(e),* Docket No. 49858, (Dec. 13, 2019).

73 *See Request for Approval of an Amended Voluntary Mitigation Plan for Luminant Energy Company LLC Pursuant to PURA § 15.023(f) and 16 TAC § 25.504(e),* Docket No. 54739 (Mar. 23, 2023); *Request for Ratification of PUCT Staff's Termination in Part of the Amended Voluntary Mitigation Plan for NRG Companies,* Docket No. 54740, Order (Mar. 23, 2023);*Request for Approval of an Amended Voluntary Mitigation Plan for Calpine Corporation Pursuant to PURA § 15.023(f) and 16 TAC § 25.504(e),* Docket 54741, Order (Mar. 23, 2023).

74 *Request for Ratification of PUCT Staff's Termination in Part of the Amended Voluntary Mitigation Plan for NRG Companies,* Docket No. 54740, NRG Notice Regarding Voluntary Mitigation Plan (Feb. 23, 2024).

75 PURA § 39.157(a).

competition would typically involve profitably raising prices materially above the competitive level for a significant period.

A key aspect in the VMPs that provided leverage in 2023 was the termination provisions. Each of the VMPs could be terminated by the Executive Director of the PUCT with three business days' notice, subject to ratification by the Commission.[76] Although the offer thresholds provided in the VMPs are intended to promote competitive market outcomes, the short-lead termination provision provides additional assurance that any unintended consequences associated with potential exercise of market power can be addressed in a timely manner.

D. Market Power Mitigation

In situations where competition is not robust and suppliers have market power, it is necessary for an independent system operator to mitigate offers to prevent the offer prices from diverging substantially from competitive levels. ERCOT's real-time market includes a mechanism to mitigate offers for resources that may have local market power because they are required to manage a transmission constraint.

Mitigation applies whether the unit is self-committed or receives a Reliability Unit Commitment (RUC) instruction. Prior to 2021, ERCOT typically issued RUC instructions to resolve transmission constraints. However, starting in summer 2021, RUCs for system-wide capacity became common and continued through early 2023. When units that receive RUC instructions are required to resolve a non-competitive transmission constraint, they often are dispatched with their offer prices capped at mitigated levels in real-time. ERCOT's dispatch software includes an automatic, two-step mitigation process:

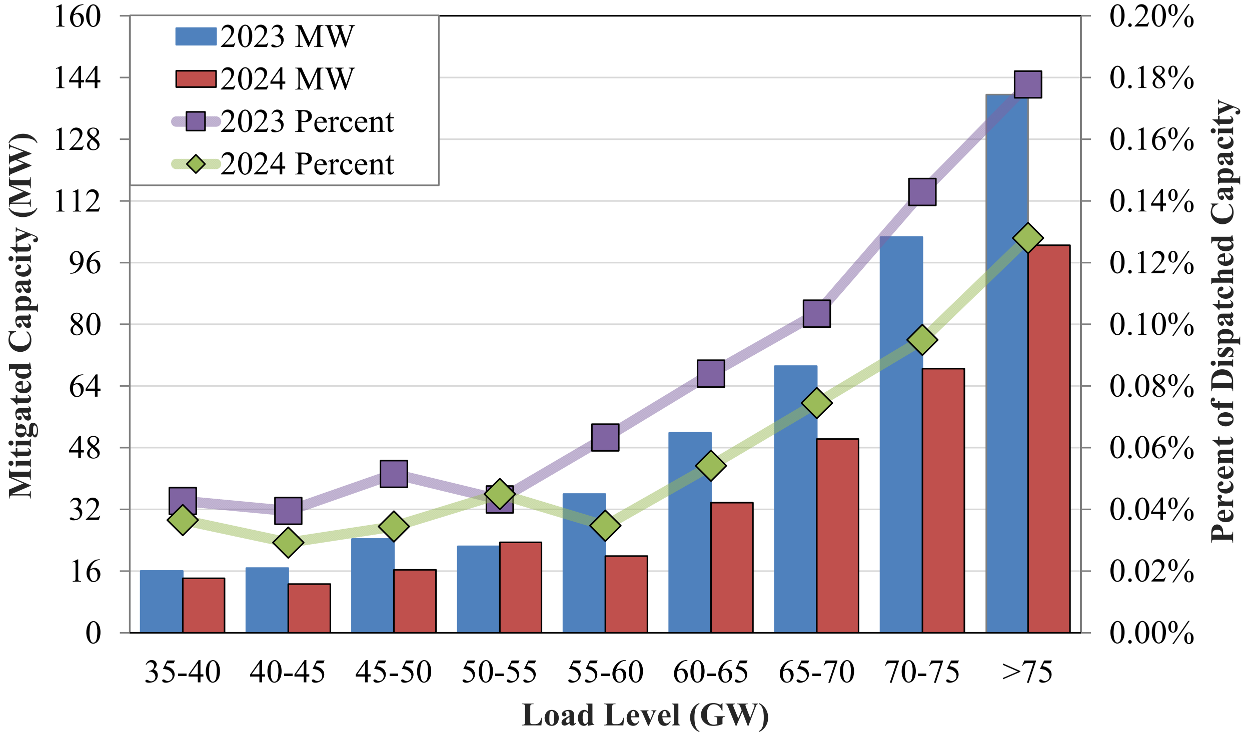
- The dispatch software calculates output levels (base points) and prices using the participants' offer curves considering only the "competitive" transmission constraints. The higher of a) resulting prices at each generator location; and b) the generator's mitigated offer cap is used to formulate the mitigated offer curve for the generator in the second step of the dispatch process.
- The dispatch software then uses the mitigated offer curve to determine the final dispatch levels and prices taking all transmission constraints into consideration.

This approach is intended to limit the ability of a generator to exercise local market power by raising its offer price to increase prices in a transmission constrained area. In this subsection, we analyze the amount of mitigation that occurred in 2024. The automatic mitigation under the two-step dispatch process only has the potential to have an effect when a non-competitive transmission constraint is active and binding in SCED. Figure 64 shows the average amount and

76 Further, Luminant's VMP will automatically terminate on the earlier of ERCOT's go-live date for Real-Time Co-Optimization (RTC), seven years after initial approval of the VMP, or the day Luminant's Installed Generation Capacity drops below five percent of the total ERCOT Installed Generation Capacity.

percentage of capacity that was mitigated at different load levels. The amount of energy that could be produced within one interval is deemed mitigated for the purposes of this analysis.

Figure 64: Mitigated Capacity by Load Level



The quantity of mitigation shown in Figure 64 is very low compared to the total quantity of capacity online. Additionally, the two-step process in ERCOT will sometimes mitigate conduct that is not significantly increasing prices and, therefore, cannot be argued to be a legitimate exercise of market power. Therefore, these results do not raise competitiveness concerns.

The extent of mitigation was less in 2024 compared to 2023. A large driver of this lies with two factors. First, 2024 exhibited less in terms of extreme weather conditions that drive higher load for shorter periods of time. Higher load levels can increase congestion which could trigger mitigation. Second, ERCOT improved its deployment of the ERCOT Contingency Reserve Service (ECRS) in 2024 which made more energy available to the real-time market, which could have reduced the incidence of congestion on non-competitive constraints and reduced the incidence of mitigation. In general, when resources are necessary to resolve a local constraint, it is more likely that the constraint will be deemed non-competitive and result in mitigation. Figure 64 also shows that mitigation tends to increase as load increases. This is also likely because higher loads can lead to more frequent non-competitive constraints binding into load pockets.

APPENDIX

Introduction

This Appendix provides supplemental analysis of certain topics raised in the main body of the Report. We present the methods and motivation for each of the analyses. However, our conclusions from these analyses and how they relate to the performance of the markets are discussed in the main body of the Report. In addition, the body of the Report includes a discussion of our recommendations to improve the design and competitiveness of the market.

I. Appendix: Statistics at a Glance

In this section of the Appendix, we provide supplemental analyses of 2024 prices and outcomes in ERCOT's real-time energy market. Table A1 is the annual aggregate costs of various ERCOT charges or payments in 2024, including ancillary services charges by type. This does not reflect the total cost of each ancillary service, as it only accounts for the net charges after self-arrangement. Also, for energy, we calculated the real-time energy value based on MWs generated rather than settlement data, as energy imbalance charges net out (plus RENA).

Table A1: ERCOT 2024 Year at a Glance (Annual)

	Annual Total ($ Millions)
Energy	$14,472
Regulation Up	$26
Regulation Down	$14
Responsive Reserve	$109
Non-Spin	$161
ECRS	$147
CRR Auction Distribution	($1,710)
Balancing Account Surplus	$239
Emergency Response Service	$73
Revenue Neutrality Uplift	$161
AS Imbalance Uplift	($8)
ERCOT Admin Fee	$292
ERO Passthrough Fee	$28
Firm Fuel Supply Service	$38
Other Load Allocation	$9
Net Cost of Electricity	$13,999

Table A2 presents the monthly aggregate costs of various ERCOT market settlement totals in 2024, including ancillary services costs by type.

Table A2: Market at a Glance Monthly

	Monthly Totals ($ Millions)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Energy	$1,531	$482	$791	$917	$1,851	$1,438	$1,110	$2,054	$1,108	$1,099	$1,123	$967
Regulation Up	$5	$1	$2	$3	$5	$1	$1	$3	$1	$2	$2	$1
Regulation Down	$4	$1	$1	$2	$2	$1	$1	$1	$0.5	$1	$0.4	$0.5
Responsive Reserve	$30	$3	$8	$11	$20	$4	$4	$12	$3	$7	$5	$2
Non-Spin	$30	$3	$10	$14	$72	$8	$2	$4	$2	$6	$5	$4
ERCOT Contingency Reserve Service	$25	$3	$8	$11	$48	$15	$5	$17	$3	$6	$3	$1
CRR Auction Distribution	($126)	($124)	($151)	($154)	($158)	($150)	($151)	($149)	($131)	($144)	($133)	($138)
Balancing Account Surplus	$40	$12	$21	$6	$19	$30	$25	$28	$17	$4	$25	$13
Emergency Response Service	-	-	-	$35	-	$3	-	-	-	$30		$4
Revenue Neutrality Uplift	$11	$8	$20	$13	$27	$9	$9	$10	$17	$9	$10	$18
AS Imbalance Uplift	$1	$0.2	$1	$2	($1)	$0.2	$0.1	($0.3)	$0.4	$1	$1	$2
ERCOT Fee	$24	$19	$20	$21	$25	$28	$28	$31	$26	$25	$21	$22
ERO Passthrough Fee	$2	$2	$2	$2	$2	$2	$2	$2	$2	$2	$2	$2
Firm Fuel Supply Service	$8	$7	$4	-	-	-	-	-	-	-	$7	$12
Other Load Allocation	$8	$1	-	($1)	$0.2	$0.1	$0	$1	$0.1	$0	$0	$0

II. Appendix: Ancillary Services

In this section, we provide supplemental data related to the provision of ancillary services through the day-ahead market and the supplemental ancillary services market (SASM).

A. Ancillary Services Provided in Real-Time

Figure A1 through Figure A5 break down the provision of each AS product by resource type. Notable trends include the following:

- Provision of RRS is dominated by ESRs and NCLRs
- The vast majority of the volume of regulation reserves is provided by ESRs
- ECRS is supplied by a combination of ESRs and gas peakers. However, duration requirements have constrained the share provided by ESRs.
- Most NSRS is provided by gas peakers. This shift began in 2022 following a sharp increase in NSRS procurement volumes, which led to greater reliance on offline units that can start within 30 minutes, primarily gas peakers.

Figure A1: Responsive Reserve Providers, 2022-2024

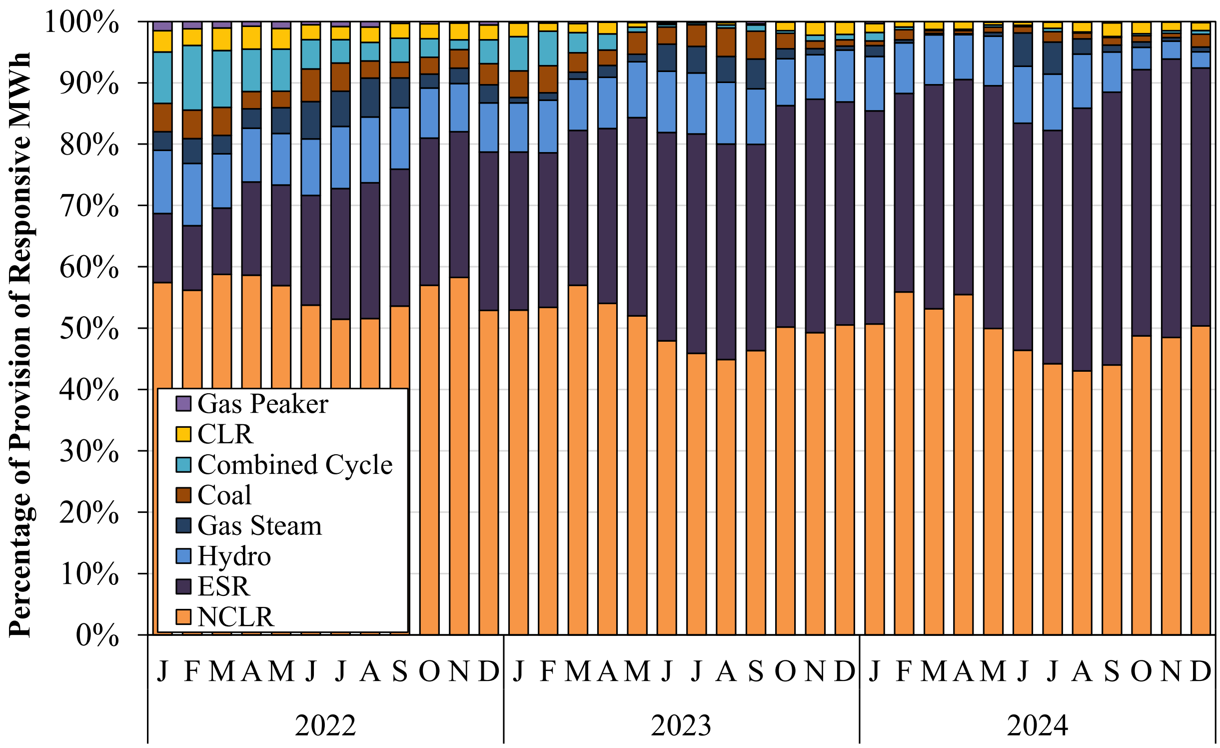


Figure A2: ERCOT Contingency Reserve Service Providers, 2022-2024

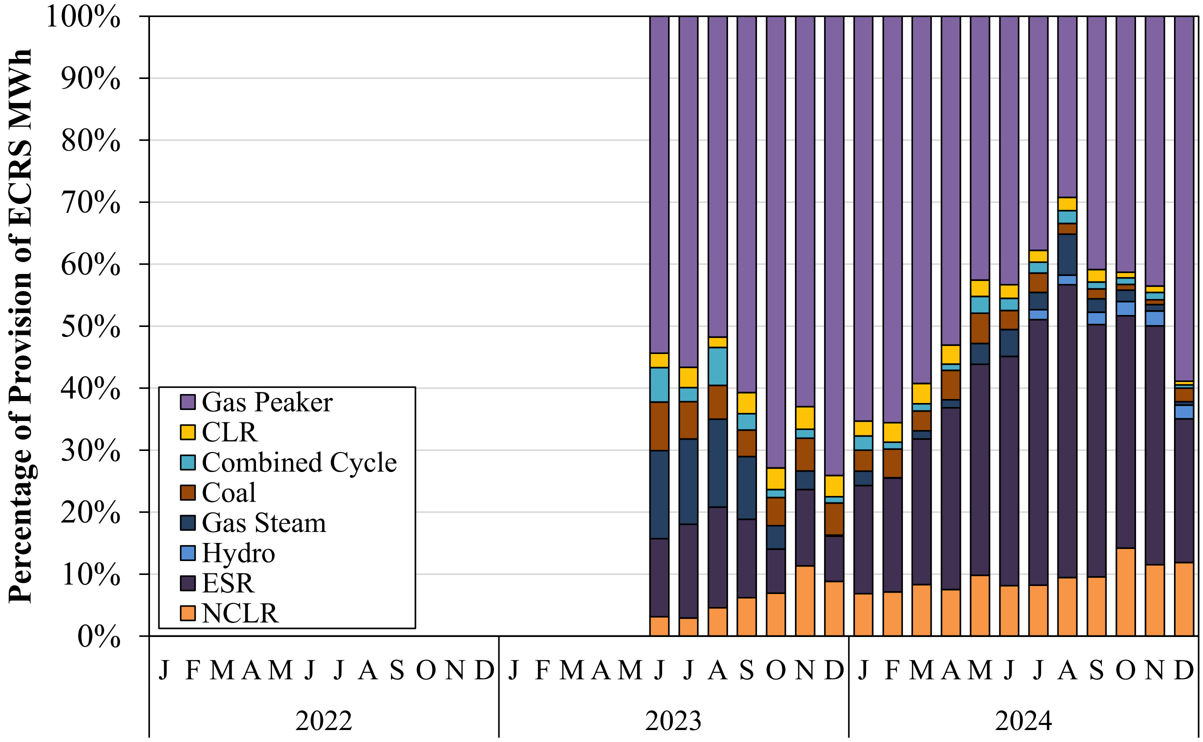


Figure A3: Non-Spinning Reserve Providers, 2022-2024

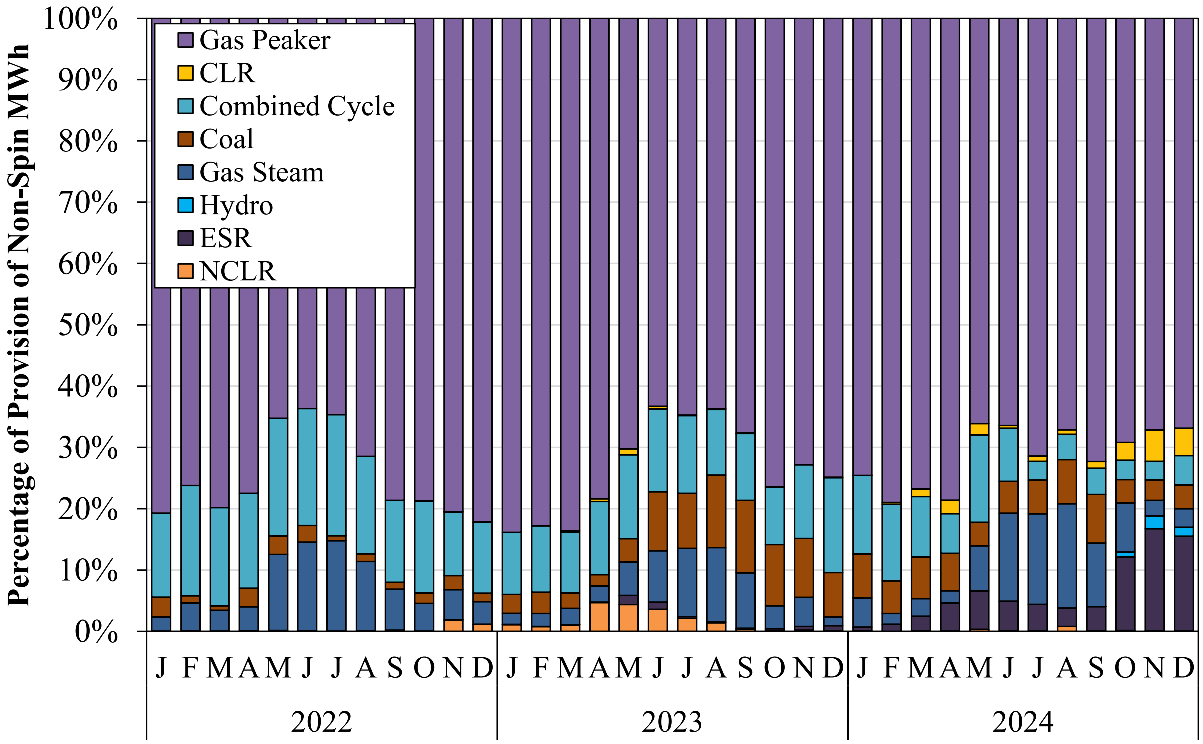


Figure A4: Regulation Up Reserve Providers, 2022-2024

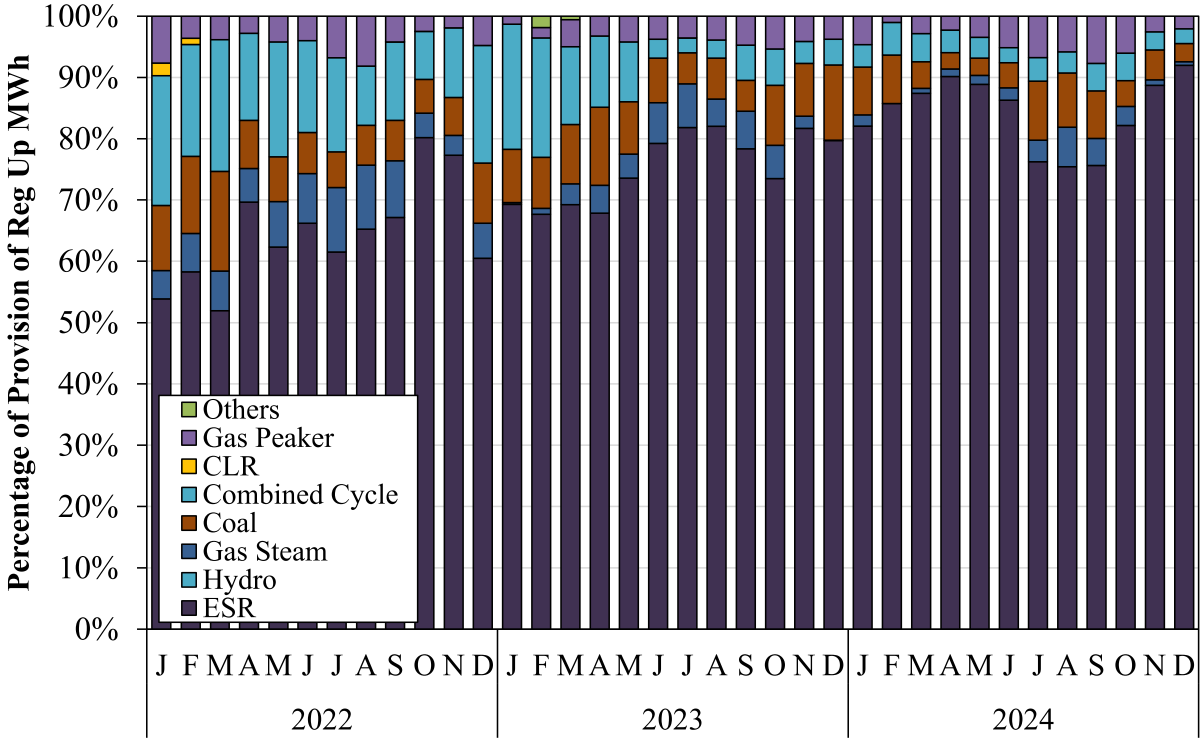
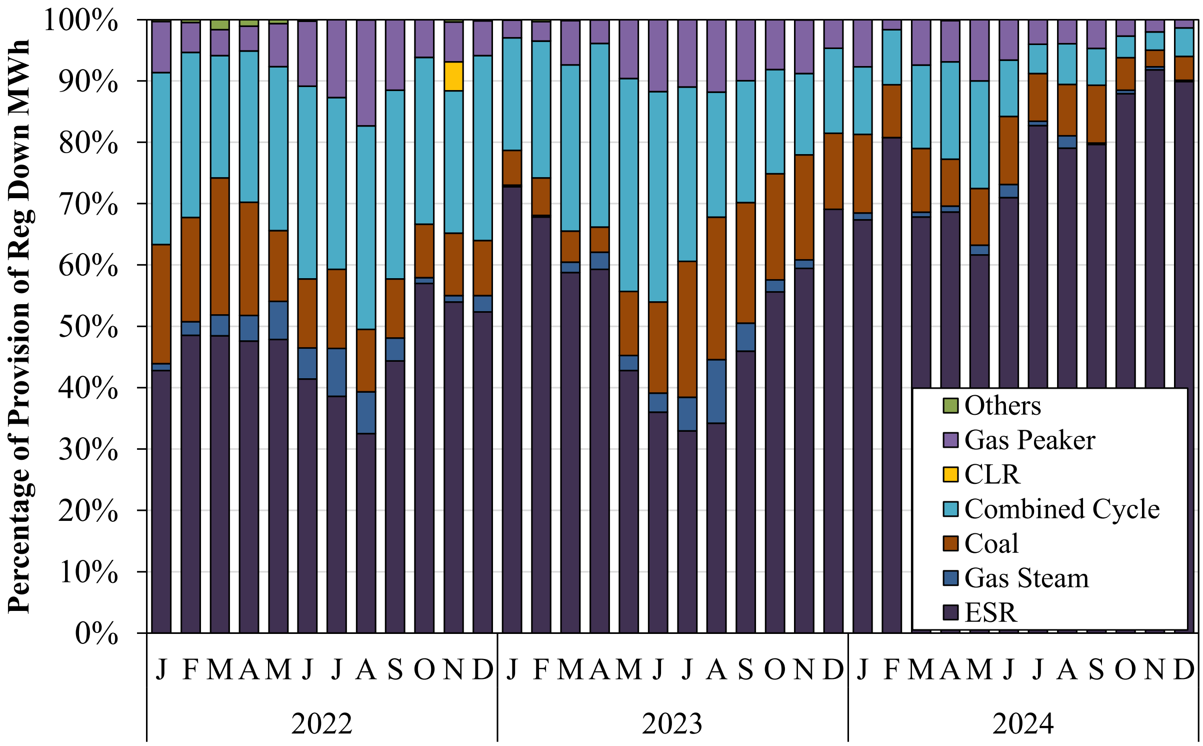


Figure A5: Regulation Down Reserve Providers, 2022-2024



B. Supplemental Ancillary Services Market

Until the implementation of RTC, the ancillary service awards from the day-ahead market are physically binding in real-time on a QSE basis. That means that if an ancillary service is awarded to a resource in the day-ahead market, the QSE for that resource can move the responsibility to carry that award to any other qualified unit in its fleet in real-time, allowing the QSE to optimize which of its resources are providing energy versus ancillary services. While these choices are likely to be in the QSE's best interest, they are not likely to lead to the most economic provision of energy and ancillary services for the whole market. Further, QSEs without large resource portfolios still face greater risks than those with larger portfolios because they may need to procure replacement ancillary services through the SASM, where prices can be high and uncertain. This replacement risk is substantial. Clearing prices for ancillary services procured in the SASM are often three to four times greater than clearing prices from the day-ahead market.

The volume of reserves procured through the SASM for 2020-2024 is shown in Figure A6. SASMs were executed 102 times in 2024 to procure a total of more than 27,000 MW of operating reserves, more than three times the volume procured through SASMs in 2023, but still very low compared to the nearly 70 million service-hours set by the AS Plan.

Figure A6: Ancillary Service Quantities Procured in SASM, 2020-2024

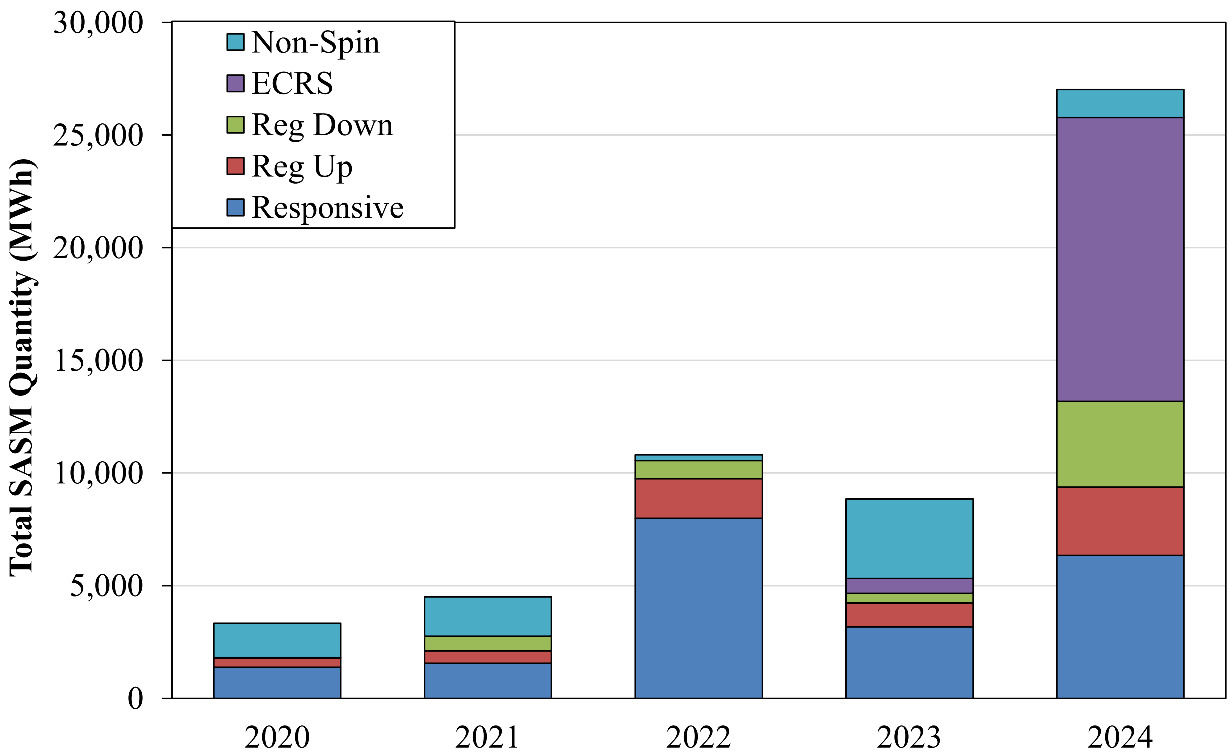
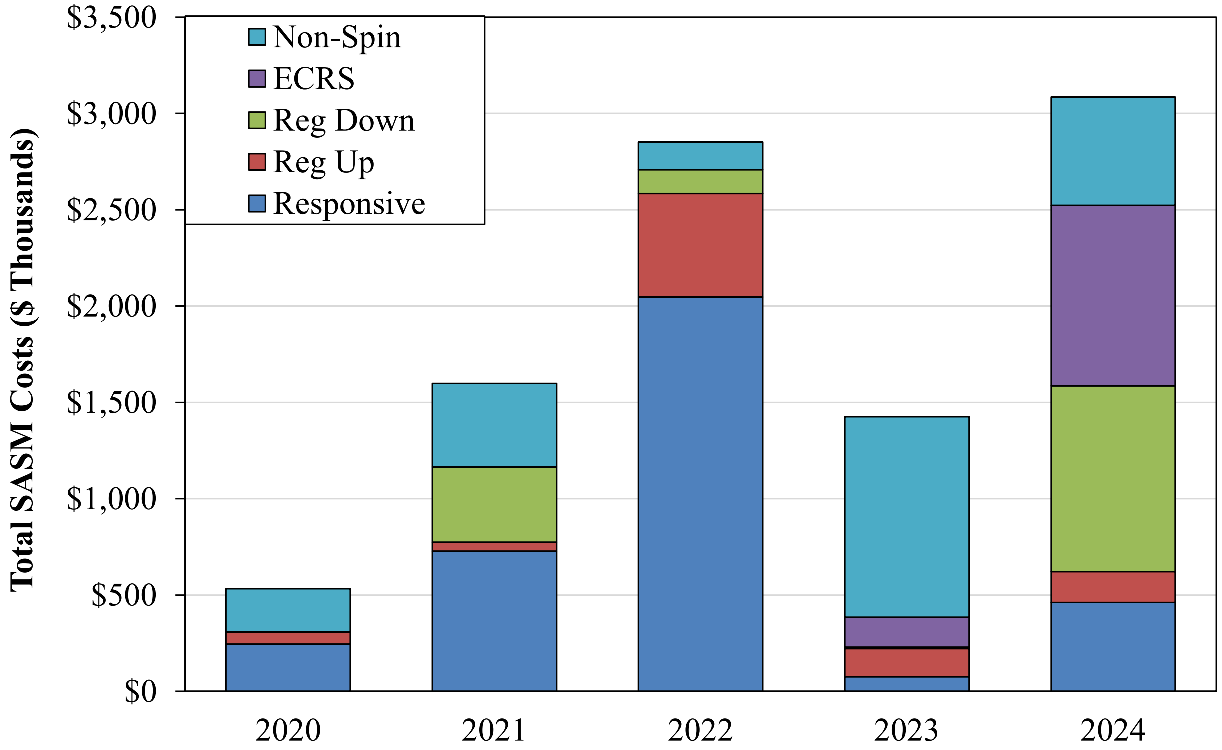


Figure A7 shows the average cost of the replacement ancillary services procured by SASM from 2020-2024. The total SASM costs across 2024 reached its peak since 2014, though only slightly higher than in 2022. SASM costs since 2021 have been substantially higher than they were from 2014-2020, which is a result of the large increases in operating reserves procured since 2021.

Figure A7: Total Cost of Procured SASM Ancillary Services, 2020-2024



III. Appendix: Detail of Existing VMPs

In 2023, three market participants had active VMPs. Each of these VMPs went through significant modifications regarding Non-Spin Reserve Service (NSRS) in March of 2023. Pursuant to those modifications, NRG's ancillary services offers are no longer covered by their VMP; Luminant has a $20 per MWh NSRS offer cap; and Calpine has a dynamic formula based on its offers for other ancillary services. NRG terminated their VMP as of March 1, 2024.[77]

i. Calpine VMP

Calpine's VMP was initially approved in March of 2013.[78] Because its generation fleet consists entirely of natural gas fueled combined cycle units, the details of the Calpine plan are somewhat different than the others. Calpine may offer up to 10% of the dispatchable capacity of its portfolio at prices up to $500 per MWh. Additionally, Calpine may offer up to 5% of the dispatchable capacity of its portfolio at prices no higher than the system-wide offer cap. When approved, the amount of capacity covered by these provisions was approximately 500 MW.

In March of 2023, Calpine's VMP was amended to eliminate the provision allowing NSRS in the day-ahead market to be made up to and including the high system-wide offer cap.[79] A dynamic formula for NSRS offers was substituted for the eliminated provision.[80] The new formula is based on Calpine's offers for other ancillary services, recognizing that NSRS are of lower value to the ERCOT system than responsive reserve service, regulation up, or ECRS. Calpine's VMP remains in effect from the date it was approved by the PUCT until terminated by the Executive Director of the PUCT or Calpine.[81]

ii. Luminant VMP

Luminant received approval from the PUCT for a new VMP in December 2019.[82] The PUCT terminated Luminant's previous VMP on April 9, 2018, as a result of its merger with Dynegy,

---

77 *Request for ratification of Commission Staff's Termination in Part of the Amended Voluntary Mitigation Plan for NRG Companies*, Docket No. 54740, (Feb. 23, 2024).

78 *Petition of Calpine Corporation for Approval of Voluntary Mitigation Plan,* Docket No. 40545, Order (Mar. 28, 2013).

79 *Request for Approval of an Amended Voluntary Mitigation Plan for Calpine Corporation Pursuant to PURA § 15.023(f) and 16 TAC § 25.504(e),* Docket No. 54741, Order (Mar. 23, 2023).

80 *Id.*

81 *Id.*

82 *PUCT Staff Request for Approval of a Voluntary Mitigation Plan for Luminant Energy Company, LLC under PURA §15.023(f) and 16 TAC §25.504(e),* Docket No. 49858, (Dec. 13, 2019).

Inc.[83] The new VMP provides for small amounts of capacity from non-quick start, non-combined cycle natural gas-fired units to be offered up to 12% of the dispatchable capacity for each unit at prices up to $500 per MWh, and up to 3% of the dispatchable capacity may be offered at prices up to and including the HCAP. When approved in late 2019, the amount of capacity covered by these provisions was less than 900 MW. In addition, the plan defines allowable limits for energy offers from Luminant's quick start combustion turbines.

Before March of 2023, Luminant's VMP provided that offers in the day-ahead market for ancillary services could be made up to and including the high system-wide offer cap. In March of 2023, Luminant's VMP was amended to place a cap on offers in the day-ahead market for NSRS of $20 per MWh for all resources.[84]

83 *See Application of Luminant Power Generation LLC, Big Brown Power Company LLC, Comanche Peak Power Company LLC, La Frontera Holdings LLC, Oak Grove Management Company LLC, and Sandow Power Company Under Section § 39.158 of the Public Utility Regulatory Act,* Docket No. 47801 (Nov. 22, 2017). On April 9, 2018, Luminant filed a letter with the PUCT terminating its VMP upon closing of the proposed transaction approved by the PUCT in Finding of Fact No. 36 of the Order in Docket No. 47801. *See also Request for Approval of a Voluntary Mitigation Plan for Luminant Companies Pursuant to PURA § 15.023(f) and P.U.C. Subst R. 25.504(e),* Docket No. 44635, Order Approving VMP Settlement (May 22, 2015).

84 *Request for Approval of an Amended Voluntary Mitigation Plan for Luminant Energy Company LLC Pursuant to PURA § 15.023(f) and 16 TAC § 25.504(e),* Docket No. 54739 (Mar. 23, 2023).