Powering Reliability Through Market Design

Addressing Rising Demand and Constrained Supply, and Stimulating Investment To Support Durable Reliability

May 6, 2026

For Public Use

Contents

Letter From David E. Mills, PJM President and CEO

Dear Stakeholders,

In January of this year, the PJM Board directed staff to undertake a holistic review of the capacity market design and investment incentives. That directive was made in recognition that the conditions that made the market work for nearly two decades have changed, and that the stress now visible in prices, reserve margins and investment pipelines reflects something more fundamental than a design that needs recalibration. The Board asked PJM to examine whether the foundational assumptions of the market remain valid – and if not, what a valid set of assumptions would require.

The paper that follows is an honest attempt at that examination. It is a diagnosis of why the current moment is structurally different from prior periods of tightness, and a framing of the choices – about risk allocation, reliability standards and market architecture – that the region must eventually make explicitly or will make by default. PJM staff has tried to be rigorous, candid and transparent about what we know and what we do not.

We have also deliberately refrained from recommending a path. Some may read that as institutional caution. It reflects something more specific: a conviction, deepened by this work, that the choices described here are not primarily engineering or design choices. They are choices about what consumers owe each other, what investors can rely on and what the "shared reliability compact" means in an era of scarcity. Those choices belong to the people and institutions with democratic accountability for their consequences – to state regulators and legislatures, to FERC, to consumers and the advocates who represent them. PJM's role is to ensure those choices are made clearly, not to substitute our judgment for theirs.

We are also conscious that this paper may not have fully mapped the design space. The three paths presented here represent the most coherent alternatives we could identify – but we do not hold them too tightly. The stakeholder community has consistently generated analysis and proposals that have advanced our collective understanding, and we expect this process to be no different. We welcome challenge, refinement and alternatives that this paper has not considered.

There is, however, one thing we want to say plainly. Wholesale electricity markets are extraordinary institutions, and their most essential infrastructure is not a price curve or a performance obligation – it is legitimacy. Generators, utilities, investors and consumers must all believe, at a basic level, that the rules are fair, stable and the product of a process they recognize as credible. That legitimacy cannot be designed in Valley Forge and delivered to the region. It is built through the kind of deliberation this paper is intended to initiate – and it is the only foundation on which a durable market design can rest.

The current situation is not tenable. The region has years, not decades, to make these choices deliberately. We are committed to that process, and we are grateful for the engagement of all who contribute to it.

Sincerely,

David Mills

President & CEO, PJM Interconnection

Executive Summary

A Region at a Crossroads

For two decades, the PJM region managed its electricity system in an era of relative stability. The Reliability Pricing Model, PJM's capacity market, was built for that environment: a system with predictable, gradually changing load; a coal-to-gas fuel transition that could be managed over a years-long horizon; and a generation development timeline that aligned with the market's three-year forward horizon. In that environment, the market worked. It provided a transparent, competitive coordination and revenue-recovery mechanism through which the retirement of more than 50 GW of aging coal and other legacy generation, and the entry of new gas-fired and other resources, could be managed and priced.

Those conditions have fundamentally changed. The PJM region is now navigating a convergence of three structural forces that have pushed the system into disequilibrium: an unprecedented surge in demand driven by the rapid expansion of large-load data centers and broader economy-wide electrification; the accelerated retirement of dispatchable generation due to environmental policy and economics; and significant supply chain and permitting frictions that have extended the time required to bring new resources online. The result is a transition from an era of managing surplus to an era of managing scarcity – one that is anticipated to persist for some time based on current projections, because new generation simply cannot be built fast enough to offset the combined effect of retiring supply and surging demand.

On Jan. 16, 2026, the PJM Board of Managers directed PJM staff to undertake a holistic review of the capacity market design and investment incentives. The Board recognized that the market's current price volatility – while economically rational – is placing unsustainable stress on the governmental compact that allows the market to function, and that the foundational assumptions of the Reliability Pricing Model design must be reexamined in a resource-constrained world. This white paper is PJM's response to that directive.

Why This Transition Is Harder Than the Last One

It is worth pausing on a question that is sometimes glossed over: Why is this structural transition more difficult than the coal-to-gas transition the markets managed successfully a decade ago?

The answer lies in a set of structural changes in the economics of generation investment – changes that are independent of any one developer's preferences and that have made the capital formation challenge qualitatively different from what it was in the last decade and up through the early 2020s.

Construction timelines have doubled. A decade ago, the market's three-year forward procurement horizon was calibrated to the development timeline of the market's reference technology. Today, a new natural gas turbine plant – the reference technology for the capacity market – requires at least four years under optimistic assumptions from financial investment decision to commercial operation, due to supply chain constraints on major equipment (gas turbines, step-up transformers), extended permitting and interconnection timelines, and larger project scale. The market's forward signal and the physical reality of construction are no longer aligned, exacerbated by delayed and compressed auction schedules since 2019.

Capital costs have escalated sharply. Installed costs for new natural gas combined cycle plants now regularly exceed $2,200 per kilowatt for projects with commercial dates in the late 2020s and early 2030s, and costs for more distant commercial dates are higher still. This represents a substantial increase over the administrative cost of new entry (CONE) estimates that PJM uses to calibrate its capacity market demand curve – and over the levels that supported investment decisions a decade ago. A 1,000 MW combined cycle plant today is a $2+ billion investment made against a backdrop of profound uncertainty about the regulatory and political environment that will prevail when the plant first generates power.

Driven by cost and uncertainty increases, revenue certainty requirements have similarly gone up. During the shale boom of the early 2010s, institutional capital was willing to finance generation projects on a largely merchant basis, accepting the uncertainty of future energy and capacity market revenues. That era is over.

The market rule changes of the past decade – Minimum Offer Price Rule expanded, then reversed; capacity accreditation methodology revised; price caps modified under governmental pressure; reserve pricing reforms accepted, then remanded; auction schedules compressed and delayed – have demonstrated, repeatedly, that the rules governing market revenues can change materially over the life of an investment. This has been compounded with significant swings in state and federal policies across administrations that impact the economic viability of generators, along with rising uncertainty in the growth of future demand. The consequence is that project finance lenders now require greater revenue certainty, typically in the form of long-term off-take agreements, to underwrite generation projects at manageable financing costs.

Outside investment alternatives have proliferated. Developers and capital that might have built merchant generation for PJM's spot market a decade ago now have access to an expanding set of alternative applications: data center construction and co-location arrangements, behind-the-meter generation serving hyperscale loads, and battery storage projects supported by tax credits. The capital seeking generation investment opportunities has not disappeared; rather, it has become more discriminating about revenue certainty. PJM's market must compete for that capital to meet our shared reliability objectives.

The combined effect is that the market design adequate to attract investment under the conditions of the prior decade is not necessarily the one that will attract investment going forward. The existing market design helped support the investment needed to maintain a high level of reliability at a low cost for almost two decades. However, the changing dynamics of the industry combined with a clear aversion to capacity price and cost volatility – a property of the current market design – indicate the need to rethink the problem and solutions. From PJM's perspective, reliability remains our absolute North Star.

Reliability remains our North Star. The resource adequacy standard the region has chosen for the last half-century requires maintaining sufficient resources on the system to meet the demand for electricity and limit the need for involuntary customer load shed to no more than once per decade.

The Credibility Trap

PJM's capacity market is currently doing what it was designed to do: signaling the value of scarce reliability by clearing at high prices in much the same way it signaled the system had excess supply by clearing at low prices over most of the last decade. But a market signal is only effective if it is durable enough to trigger the investment it is intended to prompt.

A significant portion of PJM load is served by default service providers – state-regulated entities that are often constrained by statute or regulatory practice from entering long-term bilateral contracts. These entities procure capacity primarily through short-term default service auctions calibrated to the annual Base Residual Auction cycle. During PJM's prolonged surplus era, this approach was rational: short-term procurement captured low market prices without paying a long-term hedge premium. But it also left these entities – and their retail customers – exposed to capacity price volatility when the market transitioned into scarcity. This effect was exacerbated by the persistent capacity auction delays experienced in PJM.

The result is what might be called a credibility trap: high scarcity prices that are economically necessary to signal the need for investment become, in this environment, the trigger for governmental intervention that undermines the credibility of those same prices. When prices spike and unhedged load is immediately exposed, policymakers face intense pressure to intervene through price caps, emergency backstop procurements, or other administrative actions that signal to investors that the revenue they are counting on to recover their costs may not materialize. Investors, anticipating further intervention, discount the durability of the price signal. Capital stays on the sidelines; the shortage persists.

The underlying economic logic that scarce capacity commands high prices is correct and necessary. The problem is structural: the combination of unhedged load, observed and forecasted load growth occurring faster than generation can physically be constructed and interconnected, and the perceived (and at times actual) governmental willingness to override the market's price signal prevents the market from fulfilling its investment function. Resolving this requires structural choices, not just parameter adjustments.

Foundational Questions

PJM presents the following pathways not to unilaterally prescribe a solution, but because these structural questions can no longer be deferred:

Should we preserve the concept of resource adequacy as a common good that is shared by all, also known as the "shared reliability compact" – and if so, who is responsible for making it financially durable?

For a century, PJM's grid has operated on the principle that all customers share in a common standard of reliability and willingness to pay to meet it. Preserving that compact in a period of structural scarcity requires that most load be insulated from spot market price volatility through long-term forward commitments. But mandatory hedging involves choices about who bears the burden of those requirements, how they are enforced and who holds the contracts.

Or should we decide that reliability, in a period of scarcity, must be explicitly rationed?

If not all load can be served at the historical 1-in-10 standard, the system must develop the operational capability to prioritize – whether by geography, by customer class or by contractual contribution to supply. This requires acknowledging that reliability is not always a common good shared equally by all.

And in either case, should the primary long-term hedging instrument be the capacity product, or do we make a deliberate shift of revenues supporting resource adequacy to the energy market with long-term contracting of the energy product?

If long-term contracting is required under any path, both to create revenue certainty for investors and insulate consumers from price and cost volatility, should the market be designed to incentivize long-term contracting around the separately administered capacity product or the underlying physical commodity – energy? Long-term energy contracts (power purchase agreements, tolling agreements, financial energy hedges) are more standardized, more resistant to regulatory rule changes and often more legible to project finance than bilateral capacity contracts – but a transition toward them requires reform of how Energy and Ancillary Services (E&AS) markets price scarcity.

Three Paths Forward

PJM presents three distinct paths, each reflecting a different set of answers to these foundational questions. These paths represent some of the options available for further discussion but by no means are a complete set. They are intended to initiate constructive discussion that can eventually lead to more fully developed structures.

Path A – Stabilized Markets:

Preserve the shared reliability compact by making it financially durable. Under this path, the capacity market continues, but the vast majority of load is required to be covered through long-term forward commitments either through mandatory Load Serving Entity (LSE) hedging requirements or through a PJM-administered, long-term procurement, such as a tiered, multiyear capacity market. In either case, the purpose would be to procure the vast majority of capacity needed to maintain resource adequacy prior to the final auction for the delivery period, or capacity spot market, so that it may clear at high scarcity prices when the system is short (maintaining the investment signal), but most load is insulated from those prices through forward contracts. This path trades the optionality of short-term procurement for the stability needed to support investment, and in doing so, seeks to avoid the credibility trap.

Path B – Differential Reliability:

Decide that the shared reliability compact should not be maintained for all loads in a period of structural scarcity and develop the operational and commercial framework to explicitly differentiate reliability. This could be implemented geographically (different states or zones procuring different levels of reliability) or by customer class (for example, with residential and native loads insulated from curtailment while unbacked new large load additions are curtailed first). Path B focuses on physical accountability – those who do not bring or fund supply cannot lean indefinitely on the shared pool – but it requires a fundamental reorientation of how the PJM system allocates reliability as a scarce good.

Path C – Energy Market Transition:

Pursue a deliberate, phased shift of revenue recovery from the capacity market to the Energy and Ancillary Services markets, paired with long-term forward energy contracting requirements to protect consumers from increased energy price volatility and to support investment. Like Path A, Path C assumes the shared reliability compact can be maintained; it differs not on whether to preserve universal reliability standards, but on which financial instrument – the administered capacity product or the underlying energy commodity – is best designed to sustain them. To be clear, this path should not be interpreted as a swap to an "energy-only" market. While the concept is to focus on maximizing the efficiency and transparency of the E&AS markets and making those products the central focus for valuing contributions to reliability and, ultimately, investment, we still assume the existence of a capacity market to backstop revenues needed for resource adequacy. This path explores the potential advantages and trade-offs of an explicit and progressive shift of revenue recovery from the capacity market to the E&AS markets through an increase in the E&AS market price limits (i.e., scarcity prices).

These paths are not fully mutually exclusive. Elements of Path A (mandatory hedging) are compatible with and potentially necessary for Paths B and C. E&AS market reform is a prerequisite for Path C and a beneficial complement to Paths A and B. The critical choices are about direction – which compact is the region trying to preserve, and what does that require from the market design?

The Role of the Energy and Ancillary Services Markets

In any of the three paths, PJM does not view reform of PJM's E&AS markets to be optional. Ensuring these markets efficiently commit, dispatch and price the resources and products needed to maintain operational reliability is essential. The capacity market supports the additional revenues needed to maintain resource adequacy beyond those collected in the E&AS markets. The revenue distribution between these structures, E&AS and capacity, is a choice with trade-offs that require careful evaluation. There is no right answer. This is demonstrated by the various different implementation details of each ISO's/RTO's E&AS markets.

To be effective, PJM’s E&AS markets will need to:

- **Effectively evaluate trade-offs between reliability risk and cost.**
- **Recognize operational risk and align market solutions to support operator actions.**
- **Manage changing levels of uncertainty and flexibility needs.**
- **Pre-position the system to ensure future reliability.**

If this does not happen in a least-cost fashion via the markets, and we assume the system operators take manual actions to override the markets to maintain reliability, the consequence is increased production cost due to the suboptimal, manual procurement; degraded investment and performance incentives; and unhedgeable uplift that results in cost shifts. These issues are very well known and have been discussed widely throughout the industry. Other RTOs/ISOs are ahead of PJM in these areas and are already responding to these developing demands. Most of PJM’s counterparts in other regions have undertaken or are in the process of undertaking major reserve market reforms to navigate the generation fleet’s evolution.

PJM has been working with its stakeholders in the Reserve Certainty Senior Task Force (RCSTF) to explore reforms to the E&AS markets. Within that group, PJM has put forth a proposed set of prioritized reforms to value resource flexibility and optimize dispatch based on the system reliability needs commensurate with the operational time horizon.[1] PJM considers these reforms necessary to value the essential reliability services the system needs to continue operating reliably.

The coming wave of hyperscale data center load also presents a historic opportunity. For the first time, PJM has a large and growing class of load that is technically capable of price-responsive flexibility: workload shifting, on-site backup generation dispatch and battery discharge. There are opportunities to evolve the E&AS markets to integrate these potentially participatory loads into dispatch and pricing. Successfully integrating this load flexibility into the markets would reduce the pressure on the capacity market and bridge the gap while new generation navigates interconnection queues and supply chains, while avoiding downsides of being able to access the load curtailment only through manual action during sufficiently severe emergency conditions.

PJM’s Commitment to Stakeholder Engagement

PJM presents these paths and the foundational questions that motivate them not as a unilateral market design prescription, but as the framework for a structured, transparent discussion with all relevant stakeholders. The choices embedded in these paths involve genuine trade-offs, and those trade-offs affect different stakeholders uniquely. The goal of this paper is to ensure those trade-offs are visible and that the region’s decision-making process is informed by a clear understanding of what each path requires and what it delivers.

The status quo is not a stable option. A slow erosion of reliability, masked by high and volatile prices that invite governmental intervention but fail to trigger sufficient investment, is the outcome that no stakeholder chooses when the options are laid out clearly. PJM is committed to working with all interested parties to reach a sustainable path forward – one that aligns the economic signals of the market with the reliability requirements of a 21st-century grid, and that is durable enough to attract the investment the region needs.

---

[1] PJM Interconnection, PJM’s Position on Challenges and Solutions for Long-Term Reserve Certainty Reforms (PDF), Reserve Certainty Senior Task Force, April 6, 2026, draft

I. Introduction

The region PJM serves is currently navigating one of the most complex operational and economic transitions in its history. For nearly two decades, the PJM footprint benefited from a surplus of installed capacity and relatively predictable, modest load growth. During this era, the primary challenge of the resource adequacy (RA) construct was to manage the orderly retirement of aging thermal generation while signaling the entry of competitive new resources. The market design, centered on the Reliability Pricing Model (RPM), provided a competitive mechanism through which that shift could be priced and coordinated – maintaining system security at the lowest reasonable cost to consumers. The broader coal-to-gas fuel transition also reflected structural economic forces (the shale gas revolution and declining gas costs) and environmental policy that operated independently of capacity market design, but RPM supplied the market-based framework through which those forces were translated into managed entry and exit decisions.

However, the fundamental conditions that underpinned that era of stability have shifted. The region is now navigating a rapid convergence of three structural forces: an unprecedented surge in demand driven by the electrification of the economy and the rapid expansion of data centers; the accelerated retirement of dispatchable generation due to policy and economic pressures; and significant friction in the supply chain and permitting processes that has elongated the time required to bring new resources online. The result is a transition from an era of managing surplus to an era of managing scarcity – one that is projected to persist for a decade or more, because new generation simply cannot be built fast enough to offset the combined effect of retiring supply and surging demand.

In response to these tightening conditions, the PJM Board of Managers has taken decisive action. On Jan. 16, 2026, in a letter addressing the conclusion of the Critical Issue Fast Path (CIFP) process regarding Large Load Additions, the Board accepted the immediate tactical reforms necessary to secure the grid for the upcoming delivery years. However, the Board explicitly recognized that parameter adjustments within the existing design are insufficient to address the root causes of the current strain.

The Board directed PJM staff to undertake a holistic review of the capacity market design and investment incentives. They acknowledged that the current volatility, although economically rational under the laws of supply and demand, is placing unsustainable stress on the regulatory and governmental compact that allows the market to function. The directive is clear: PJM must evaluate whether the foundational assumptions of the RPM design remain valid in a resource-constrained world and explore structural reforms that can ensure reliability is not just physically available, but economically and politically durable.

Scope and Approach

This paper is the direct response to that directive. Its purpose is to diagnose the structural friction between the current market design and the evolving physical reality of the grid and to frame the decision points necessary to resolve it. This paper does not propose detailed market design solutions. It does, however, seek to initiate discussion on the opportunity to make fundamental changes to the structure of the markets in a manner that acknowledges those structural frictions challenging it today.

To understand where we can go, we must first understand how we arrived here.

This paper will first provide background on the origins of the PJM capacity market, explaining the “missing money” problem it was designed to solve and the “shared reliability” compact it was intended to uphold.

Next, the paper will provide a candid assessment of the current resource adequacy landscape. It will examine the primary drivers of the current shortage, distinguishing between physical constraints (which markets can signal but not immediately solve) and market design gaps (which exacerbate the pain of the shortage). Specifically, the paper will analyze why the existing market is straining to support the necessary investment. It will explore the concept of the "reliability externality," where rapid load growth leans on the system's shared reserves, and the challenge of unhedged volatility, where price signals intended to spur investment instead trigger affordability crises for native load.

Crucially, this paper asserts that the market cannot solve physical constraints overnight, but the design of the market determines how those constraints are managed and how quickly they are resolved. It also discusses why the era of purely merchant generation development is largely over – not because developers have changed their preferences, but because the investment environment has changed structurally. Capital costs for new generation have escalated sharply, and construction timelines have doubled. Additionally, the compounding uncertainty of market rule changes over the past decade, swings in state and federal energy policies, and large load growth uncertainty have hampered the ability of capital investment to finance projects based on projected market revenues. Project finance lenders have responded by requiring long-term off-take agreements or equivalent revenue certainty to underwrite generation at manageable financing costs. This is not a temporary market condition; it reflects a genuine structural shift in what the investment community requires to commit capital to large, long-lived generation assets in the PJM footprint.

Crucially, this paper asserts that the market cannot solve physical constraints overnight, but the design of the market determines how those constraints are managed and how quickly they are resolved.

It's also helpful to acknowledge that the market alone – even if perfectly designed – cannot always ensure that resource adequacy is met.

There are many factors beyond the market design that impact resource adequacy on the system and the ability of the market to achieve it – a number of which are explored in this paper as contributing factors of the current shortfall.

This paper is not designed to unilaterally dictate a solution, but to provide a framework to develop and assess alternative designs for the PJM resource adequacy framework. By laying out the trade-offs between these paths, PJM aims to facilitate a transparent discussion with stakeholders, states and the PJM Board to determine a sustainable path forward, one that aligns the economic signals of the market with the reliability requirements of the 21st-century grid.

II. Background: The Missing Money Problem and the Origins of RPM

▸ **Key Takeaways From This Chapter:**

- Electricity is unlike almost any other product: Most customers pay flat rates and never see real-time wholesale prices, so they have no reason to reduce consumption when supply is tight.
- Compounding this, the grid generally cannot direct power only to customers who planned ahead – physics determines where electricity flows, which means everyone shares a shortage equally, regardless of who arranged supply.
- These two features combine with energy price caps to create the missing money problem: Generators that run only during rare, high-stress hours cannot earn enough from those hours alone to justify being built.
- PJM's capacity market (RPM) was the region's solution: a competitive, pooled procurement in which PJM acts as a purchasing agent for all customers to provide the missing money and to share the cost across the entire grid.
- The track record was strong: 50+ GW of aging resources retired and were replaced with 40+ GW of new generation, largely built with private capital rather than ratepayer guarantees.
- But the market was designed for specific conditions: slow, predictable load growth; generation buildable in roughly three years; and investors willing to accept annual auction revenues as sufficient return.
- All three of those conditions have now changed.

As PJM undertakes this review, it is imperative to state clearly: PJM remains steadfast in its support of competitive wholesale electricity markets.

The transition from vertically integrated monopolies to competitive markets was not merely a regulatory exercise; it was a fundamental shift designed to drive efficiency, innovation and cost discipline for the benefit of the consumer.

The Federal Energy Regulatory Commission (FERC) has long held that competition is the most effective mechanism to ensure "just and reasonable rates." By requiring generation resources to compete to serve load, the market ensures that the system's energy needs are met by the most efficient resources available, minute by minute. This saves consumers billions of dollars annually by optimizing the use of the entire regional fleet rather than fragmenting operations across smaller utility silos.

Furthermore, competitive markets introduced a critical shift in risk allocation. Under the traditional rate-of-return regulation model, the risks of investment decisions (e.g., cost overruns, technology failures or uneconomic construction) were largely borne by captive ratepayers. In the PJM market design, that investment risk is shifted to the investor. If a generator builds a plant that is too expensive or technologically obsolete, the shareholder, not the ratepayer, bears the loss.

However, the efficacy of a competitive market relies on its ability to send accurate price signals that reflect the true state of the system. When those signals align with physical reality, the market effectively coordinates entry and exit. It is only when the physical reality diverges from the assumptions driving the market's design, as we see in the current scarcity environment, that we must intervene to recalibrate the market and realign on foundational objectives. The goal in this paper is not to abandon the market, but to ensure the market can continue to value and deliver reliability in an evolving grid.

The Resource Adequacy Problem in Electricity Markets

The first question to ask to understand why the current system is under strain is this: *Why is "resource adequacy" a problem that needs to be solved in electricity markets but not others?*

There is no "car adequacy market" or a "bread adequacy market." In those industries, if supply runs short, prices rise. Consumers, seeing the high price, voluntarily buy less (i.e., demand response). If the shelves go empty, the store closes for the day. The market clears, albeit painfully, without catastrophic failure. Path B, as discussed in the Executive Summary, is just that. Those willing to pay the cost necessary to ensure resource adequacy for their region are rewarded with the benefits of that. Those that are not will experience a lower level of reliability but the benefit of lower costs.

Electricity is unique because it defies this standard market logic due to two structural characteristics that were taken as immutable facts when these markets were designed two decades ago.[2]

Structural Characteristic 1: Inelasticity of Demand

The first structural characteristic is an almost complete disconnection between the price of producing electricity and the price paid to consume it.

In a functioning market – whether for airline tickets, ride-shares or natural gas – price serves as a communication tool. When supply is tight, prices rise. This signal travels instantly to the consumer, who then makes an economic choice: "Is this trip or this product worth the premium price?"

If the answer is "no," they voluntarily reduce their demand and the market "clears" – that is, there is a price at which the quantity that sellers wish to sell equals the quantity that buyers wish to buy.

In the electricity sector, however, this signal is severed at the wholesale-retail interface. While the wholesale energy spot market may signal extreme scarcity with prices spiking to $3,000/MWh or higher, the vast majority of end-use customers – particularly residential and small commercial loads – are insulated from this reality. They operate under retail tariffs or contracts that typically insulate consumers from real-time wholesale price volatility. Although some tariffs include adjustment clauses that pass through capacity and energy costs over time, those adjustments are lagged and averaged rather than real-time signals.

Even when the energy market may signal extreme scarcity, the vast majority of end-use customers are insulated from this reality.

This is very different from a forward hedging mechanism where a customer, or their Load Serving Entity, purchases a defined quantity of energy at a fixed price. In that case, the customer pays the forward price for the fixed quantity, and either liquidates the unused energy at the spot price or pays the spot price for additional consumption. In the case of fixed retail tariffs, end-use customers are eligible to consume as much or as little as they desire at the fixed tariff rate.

---

[2] The discussion in the subsections that follow is adapted from Steven Stoft. Power System Economics: Designing Markets for Electricity (2002). IEEE Press Wiley.

See also similar presentations of the resource adequacy problem in:

Frank A. Wolak. "Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables," *Environmental and Energy Policy and the Economy*, vol. 3 (University of Chicago Press, 2022), pp. 155–220.

Severin Borenstein, James Bushnell, and Erin Mansur. "The economics of electricity reliability." Journal of Economic Perspectives 37.4 (2023), pp. 181–206.

This creates a perverse incentive structure. During a grid emergency, the marginal cost of consuming an additional kilowatt-hour may be one hundred times higher than the average cost. Yet, the consumer sees little or no immediate change in their bill. Consequently, they have no financial incentive to reduce consumption. They do not adjust their thermostats or delay running appliances because the price they face does not reflect the crisis on the grid.

Crucially, this is not merely a technological problem (i.e., "consumers don't have smart meters"), it is an economic one. Even with smart meters, if the customer is on a fixed-rate plan, a smart meter is simply a more precise way to measure unresponsive demand. Because customers are largely shielded from volatility, they have little economic reason to invest in the capabilities (e.g., smart thermostats, home batteries or automated load control) that would allow them to be responsive. The flat-rate structure effectively kills the business case for retail demand-side flexibility in response to prices and, in general, leads to the need to develop various demand curves (e.g., Variable Resource Requirement Curve, Operating Reserve Demand Curve) used in electricity market design as a way to mimic demand participation.

For the system operator, this results in a demand curve that is effectively vertical (nearly perfectly inelastic). With limited exceptions, the quantity demanded is effectively unchanged as the price rises. Because the market cannot rely on clearing through price response, the system operator must rely on always having available supply-side reserves to balance the system, necessitating administrative rules and "missing money" mechanisms.

Structural Characteristic 2: Non-Excludability and the Impossibility of Enforcing Physical Contracts

The second structural characteristic is the inability of the system operator to physically enforce contracts during a shortage.[3]

In a functional commodity market, like the market for wheat or flour, contracts act as a physical guarantee. If Bakery A contracts flour deliveries six months in advance and Bakery B chooses to rely on whatever was available on store shelves each day (the daily spot market), a shortage reveals a clear distinction. When flour runs out, the supplier delivers to Bakery A. Bakery B, having failed to secure supply, goes empty-handed. The market naturally enforces the contract: The entity that paid for the product or service receives it while the entity that did not or could not, does not.

Electricity markets, however, were designed under the assumption that this type of physical differentiation is impossible. The grid was designed to operate as a "common pool" resource. Once electricity is injected into the transmission system, it flows according to the laws of physics (Kirchhoff's laws), not the laws of contract. PJM cannot tag a specific megawatt and ensure it flows only to the customer who paid for it. This creates a profound enforcement problem during scarcity. If the system is short of generation, the grid operator generally cannot selectively curtail only those LSEs or those specific customers who did not arrange for supply to prevent them from nevertheless drawing from the grid when there is a supply shortage. The electricity markets were designed under an assumption that we lack the technological granularity to say: "The data center on this feeder didn't buy capacity, so turn them off, but keep the hospital next door running because they signed a long-term contract."

> The grid operates as a "common pool" resource. Once electricity is injected into the transmission system, it flows according to the laws of physics.

Instead, when the system faces a shortfall, PJM must implement manual operational protocols to involuntarily shed firm load (e.g., rotating outages). When placed in comparison to the granularity of nodal pricing and associated

---

[3] As discussed in Section IV, technological advances and the rise of a new type of customer are beginning to challenge this assumption, opening opportunities for the markets designed around this assumption to evolve.

incentives, these protocols are somewhat blunt. They cut power to entire substations or geographic zones based on engineering necessity, without regard for the commercial arrangements of the customers within those zones. The physics and engineering of the bulk power system historically have not allowed precision beyond a certain level.

Consequently, the “prudent” load that bought capacity effectively shares its reliability with the “imprudent” load that did not. In a load shed event, the entity that planned ahead may be blacked out just as surely as the entity that leaned on the system. This inability to exclude non-payers from the benefits of the grid turns reliability into a common good. It creates an inherent incentive to “free ride,” knowing that if the ship sinks, it sinks for everyone, regardless of who bought a ticket.

The Consequence: The “Missing Money” Dilemma

The convergence of these two structural limitations – inelastic demand and the inability to exclude non-payers – creates a critical market failure. Because the demand side cannot practically respond to price signals to clear the market during shortages, the system cannot rely on price alone to balance supply and demand without risking catastrophic failure. This creates a shared adequacy problem that electricity markets alone cannot solve – *determining the amount of capacity that provides an optimal or acceptable level of system reliability to avoid periods of involuntarily firm load shed or blackouts* – and requires administrative rules.

In an energy-only market design, the price cap during scarcity conditions could be based on an administrative estimate of Value of Lost Load (VOLL), a theoretical ceiling that likely reaches many tens of thousands of dollars or more. While from one perspective such pricing is efficient in that it reflects a true estimate of the marginal value of reliability, it is often viewed as unsustainable by government and practically dangerous. To prevent the exercise of market power during inelastic periods and to protect consumers from extreme volatility, regulators and system operators have historically imposed administrative price caps well below VOLL, understating the cost of load shed during shortage by an order of magnitude or more.

While lower price caps in the energy market protect participants in the immediate term, they create a long-term financing gap known as the “missing money” problem. A resource, particularly a peaking unit that runs infrequently, might rely on those few hours of extreme scarcity pricing to recover its fixed capital costs (the Cost of New Entry, or CONE). When the market design caps the price at a significantly lower level, it effectively precludes the revenue the generator requires to justify the investment. Without a mechanism to recover this “missing” capital cost, the necessary resources are simply not built, reliability degrades and the system is unable to meet the historical standard of one firm load shed event per decade.

The Current Solution: The Capacity Market as a Regulatory Proxy

When the Reliability Pricing Model (RPM) was originally designed, these structural characteristics were accepted as immutable constraints of the electricity sector. To resolve the resulting revenue gap, PJM and other ISOs/RTOs established capacity markets to ensure all loads that benefited from shared system reliability also contributed their fair share to help support it.

In PJM’s implementation, PJM acts as the centralized purchasing agent on behalf of all LSEs. Notwithstanding regions participating in the Fixed Resource Requirement (FRR) Alternative, PJM performs this function for the entire system in aggregate, enforcing a shared level of reliability across all customers. By defining a target reliability standard – historically the “1-in-10” standard, which targets a loss of load expectation of one day in ten years – PJM effectively introduces a demand-side proxy for the willingness to pay for reliability that society would purchase at if it could coordinate toward an efficient outcome in the absence of the demand-side flaws (the Variable Resource

Requirement Curve). Additionally, the implementation allows demand response to participate and provide their own willingness to pay for capacity.

Through the forward auction mechanism, PJM procures a quantity of capacity consistent with the supply offered and demand's willingness to pay three years in advance, providing a transparent, competitive price signal to generators. The resulting capacity payment provides the opportunity to recover the "missing money" required to cover the fixed costs for those resources receiving an award. In exchange for this payment, the generator effectively sells a physical call option to the region's consumers. They accept a rigorous obligation to perform during system stress events, effectively pre-selling their availability to the pool. This architecture is the institutional foundation of what this paper calls the shared reliability compact: a pooled procurement design that addresses the non-excludability problem by ensuring everyone who benefits also contributes to a common standard.

This construct has provided the organizational and financial framework through which PJM maintained resource adequacy for two decades – a period during which underlying market conditions (abundant new entry, stable demand growth, relatively short construction timelines) were broadly favorable to its design assumptions. However, it was built on a foundational assumption of homogeneity: that all customers, with the exception of those voluntarily participating in demand response programs, were equally passive, and that reliability was strictly a common good that must be procured centrally for everyone.

As we look to the future, and as we discuss later in this paper, we must ask: Are these assumptions still valid? Or does the rise of more sophisticated large loads and evolving policy preferences allow, or possibly require, us to rethink certain foundational assumptions that lead to a one-size-fits-most approach?

History of PJM's Capacity Construct

PJM's capacity construct has evolved over time into the market design in place today. Appendix A provides a historical overview of that evolution and reviews the predecessor capacity constructs and drivers that led to the current market framework (the Reliability Pricing Model or "RPM"), the primary design elements when RPM was implemented in 2007, certain notable modifications to the market design over the past two decades, and some of the key successes of the market since its inception.

A foundational element that has remained throughout its evolution is the inherent value of the "PJM pool." PJM members, and ultimately the consumers across the footprint, have long benefited through the sharing of available resources to reliably serve load. The aggregation of resources and the diversity in regional load profiles across PJM's large geographic footprint helps mitigate risk, and it allows the system to reliably serve loads with a much lower reserve margin than what any single service territory would require in isolation. This dynamic has provided significant reliability value and savings for customers in the PJM footprint over the years.

PJM has also long relied on resource adequacy standards and a capacity construct, or mechanism that sets capacity requirements on LSEs, to support reliable service to loads. Since the 1960s, PJM has used probabilistic methods to determine the pool-wide reserve margin and capacity needed to meet the one day in 10 years LOLE standard of the system. The analysis accounts for uncertainty in factors affecting reliability, such as generator performance and load characteristics, to determine the total installed capacity needed to satisfy the region's LOLE criterion. This targeted capacity level is expressed in terms of a percent above peak load (Installed Reserve Margin or IRM) and provides the baseline for setting the capacity requirements of LSEs in the pool. In addition to the pool-wide resource adequacy standard, PJM has long used deliverability standards to test the transmission system's ability to deliver energy from, and to, various parts of the footprint to help ensure locational reliability.

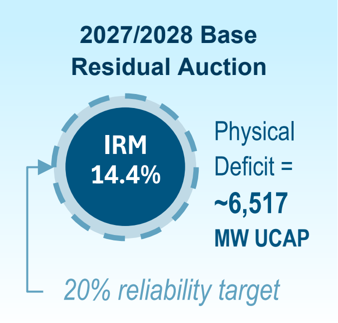
III. Current Resource Adequacy Landscape

▶ **Key Takeaways From This Chapter:**

- PJM is facing a period of structural scarcity, a fundamental mismatch between how fast demand is growing and how fast new supply can follow.
- The region simultaneously faces hyperscale data centers adding load at an unprecedented pace, accelerated policy- and economically-driven retirements of generation, and new power plants that now take roughly twice as long to build and cost twice as much as they did a decade ago.
- New large loads are connecting faster than the fleet can expand to serve them, drawing on a shared reliability reserve that existing customers have paid to maintain – before contributing commensurate supply themselves.
- High capacity prices are the market's rational response to these conditions, but those prices translate directly into customer bills for a significant portion of load, particularly in restructured states, due to a lack of long-term supply contracts and hedging to buffer against the rapid rise in the wholesale's spot price.
- These pressures create a credibility trap: Scarcity prices trigger affordability concerns and political pressure to intervene; investors see that pressure, discount the future revenue signal and hold back investment — so the shortage persists.

While the capacity market helped successfully manage the transition from surplus to equilibrium over the last decade, the current environment of rapid disequilibrium is stress-testing the market's foundational logic. The strain we observe today is not simply a matter of "tight supply," it is the symptom of a deeper structural friction. The implicit social and economic contracts that underpinned the RPM design, specifically the "shared reliability compact" and the "investment signal," are eroding under the weight of new realities.

The results of the 2027/2028 Base Residual Auction provide the concrete backdrop for the structural analysis that follows. The auction cleared approximately 134,478 MW of Unforced Capacity (UCAP), yielding an Installed Reserve Margin of 14.4% – a 5.6 percentage point shortfall against PJM's 20% reliability target and a physical deficit of approximately 6,517 MW UCAP. As required under the Tariff, PJM conducted an investigation into the root causes of the capacity shortfall and published its findings in a report.[4] That report recognized several contributing factors, but the primary driver was the surging demand and data center load growth outstripping the pace at which infrastructure can be built to supply it. This shortfall is only projected to grow over the next decade as new large loads come on to the system.



---

[4] PJM Interconnection, 2027/2028 Base Residual Auction Reserve Target Shortfall Report (PDF) (2025)

Diagnosing Scarcity: Market Mechanics vs. Structural Inevitability

As the region grapples with the sudden transition from surplus to shortage, it is necessary to separate the symptoms of this transition from its root structural causes. A common critique is that the recent compression and delay of PJM's Base Residual Auctions (BRAs) manufactured the current affordability crisis. We must address this explicitly: To what extent are delayed auctions responsible for the problems we face?

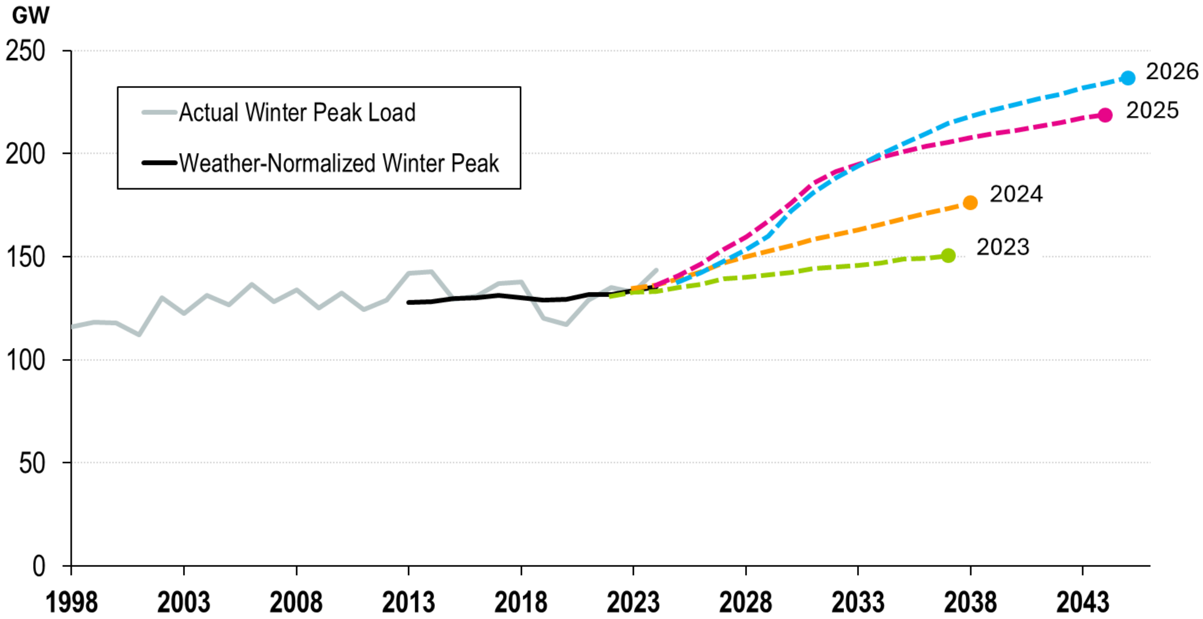
The delayed BRAs undeniably exacerbated the price shock of the shortage. By falling out of sync with the default service auction timing of many retail states, the compressed schedule bypassed the mitigating effects of laddered procurement built into several restructured (retail choice) states' default service auctions for customers choosing to receive supply from the default utility or "provider of last resort." Instead of a gradual, multiyear price increase that signaled tightening supply, unhedged loads were hit with a sudden, jarring exposure as capacity clearing prices jumped an order of magnitude. This immediate affordability crisis triggered governmental intervention and regulatory uncertainty, as we further explore in the subsection titled "The Gap Between Price Signals and Durability With Government."

However, while the auction schedule dictated how the price shock was delivered, it did not create the underlying shortage, nor could the regular auction schedule have stopped it. The scarcity we are experiencing is driven by exogenous macroeconomic trends and state and federal policy decisions that outstrip market mechanics. Even if the auctions had remained perfectly on schedule, blunting the immediate sticker shock, the standard one- to three-year forward hedging utilized by most default service providers is wholly insufficient to protect consumers from the cost impacts of a sustained supply deficit. In short, load is being interconnected far faster than generation can physically be built to match it; the system is going short, and the market is pricing the shortfall.

The Structural Mismatch: Was This Avoidable?

We must acknowledge that this disequilibrium was largely inevitable. Regardless of the forward period of the market, state and federal policies over the last decade combined with low capacity prices to encourage or require the retirement of generation faster than its reliability value was replaced. A slow-moving supply chain best suited for the relative stability of the last several decades further exacerbated the impact of an unprecedented, continent-wide surge in macro-load. **Figure 1** shows the rate of forecasted growth of peak load in winter, the season in which PJM currently observes the majority of resource adequacy risk; forecasted summer peak load growth shows a similar pattern. Despite the complexities of the capacity market, the resource adequacy problem is quite simple: load is forecasted to grow at the fastest rate in almost half a century, and supply cannot expand fast enough.

Figure 1. Historical and Projected Winter Peak Load Growth in PJM



Crucially, a foundational assumption of the RPM design has been physically broken. The three-year forward timeline was originally established because it aligned with the construction schedule of many of the resources that could be built to meet the system's resource adequacy needs, including both combustion turbine and combined cycle gas resources. It was assumed, and largely true, that a high price signal in year zero, combined with reasonable expectations of future market revenues, could yield physical steel in the ground by year three. Today, due to compounding frictions in siting, permitting, supply chains and interconnection, it can take two or three times as long to build a gas resource as it does to build a hyperscale data center. These hurdles are examined in further detail in the following section of the report.

Therefore, even if all regulatory and market processes had operated flawlessly, the sheer velocity of data center load growth combined with a sluggish supply-side response significantly increased cost, and magnified regulatory risk of generation development guaranteed a severe physical mismatch. This is further corroborated by similarly stressed grid conditions in regions across the United States.[5]

PJM's generation interconnection queue could be viewed as a once-binding constraint. However, PJM's interconnection reforms have made substantial progress in clearing the administrative backlog: Approximately 57 GW of projects have completed interconnection studies and have been offered or executed Generation Interconnection Agreements, including 14 GW in Transition Cycle 1 (84 projects, predominantly solar and storage) and nearly 30 GW in Transition Cycle 2 (278 projects, including 31 natural gas projects over 5 GW).[6] The pipeline is real. The challenge is conversion. Since 2020, approximately 24 GW of projects with fully executed interconnection agreements have terminated before reaching commercial operation – including 13.5 GW of natural gas projects. These were not

---

[5] NERC 2025 Long Term Reliability Assessment (PDF) report issued in January 2026: "*The overall resource adequacy outlook for the North American BPS is worsening: In the 2025 LTRA, NERC finds that 13 of 23 assessment areas face resource adequacy challenges over the next 10 years. Projections for resource and transmission growth lag what is needed to support new data centers and other large loads that drive escalating demand forecasts*."

[6] PJM Interconnection, 2027/2028 Base Residual Auction Reserve Target Shortfall Report (PDF) (2025)

projects that never existed; they were projects that cleared the queue, secured their interconnection rights and then failed to convert due to permitting denials, supply chain delays or financing that could not be closed.

These structural forces are sufficiently strong enough that it appears unlikely that PJM is experiencing a brief, one-or-two-year transition. We will not simply “build our way out” of this deficit in the near term. We are facing a possible decade-long structural reality where demand growth will continually threaten to outpace supply additions.

What Does It Cost To Build a Generator in PJM?

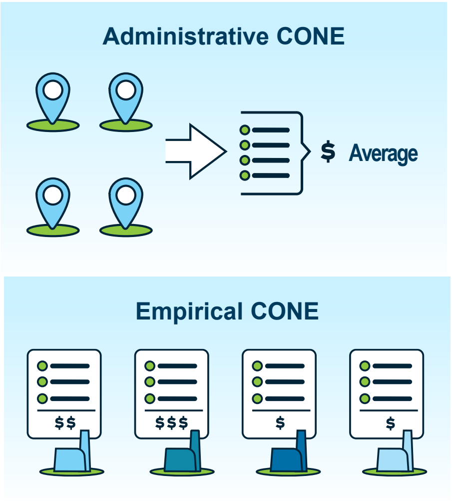
Understanding the investment challenge requires an understanding what new generation actually costs today – and how dramatically those costs have changed in a very short period of time.

Administrative CONE vs. Empirical CONE

In a well-designed capacity market, the Cost of New Entry study should answer a straightforward question: what does a new generator need to recover to make investment viable? As the anchor of the VRR curve, CONE serves as the market’s reference point for long-run entry economics — the price level around which clearing prices should orbit to maintain a reliable system over time. In this sense, there is conceptually only one CONE: the administrative estimate produced by engineering studies and the empirical reality of what developers actually pay should, in long-run equilibrium, be the same thing.

They are not the same today, and the gap is not incidental. Administrative CONE studies are engineered to produce a stable, long-run average — representative of normalized market conditions rather than today’s spot realities, and deliberately smoothed to insulate the demand curve from short-run volatility. What developers actually pay reflects current supply chains, current financing markets, and the risk premium that lenders require given real-world revenue uncertainty. When current conditions depart significantly from long-run equilibrium, as they have, the administrative estimate will systematically understate the market-clearing price needed to attract new investment. Understanding both numbers, and the direction of the gap between them, is essential to diagnosing whether the capacity market is sending adequate investment signals for the reliability challenge the region now faces.

PJM’s capacity market uses a periodic Cost of New Entry (CONE) study to calibrate the demand curve that governs capacity market pricing. It is important to distinguish between two related but distinct cost concepts:



Administrative CONE is the bottom-up estimate of the capital and fixed costs to build a reference technology (historically a combustion turbine, and in recent reviews also a combined cycle plant and battery storage system) in representative locations across the PJM footprint. This is the number produced by studies like PJM’s Sixth Quadrennial Review, conducted by The Brattle Group and Sargent & Lundy and submitted to FERC in 2025. It is a carefully constructed average across multiple sites and regions, intended to represent a stable, long-run view of new entry economics rather than a snapshot of today’s elevated market conditions.

Empirical CONE is what developers actually pay to build a specific plant at a specific site – and it is the number that determines whether investors will commit capital. It is derived from actual developer filings: Certificates of Public Convenience and Necessity (CPCNs), Integrated Resource Plans (IRPs), project finance disclosures and original equipment manufacturer (OEM) pricing. These figures capture site-specific costs (land, permitting, interconnection, pipeline), current supply chain conditions and the financing premium that lenders require given today's revenue uncertainty. They are project-specific and not universally representative – but they represent the actual investment decisions that the market needs to trigger.

The gap between these two numbers has widened dramatically, and the direction of that gap matters enormously for both the capacity market's ability to attract investment and for consumers' long-run cost exposure.

The 2025 Brattle CONE Study: A Market at an Inflection Point

PJM's Sixth Quadrennial Review, conducted by The Brattle Group and Sargent & Lundy (2025), provides the most comprehensive and current administrative assessment of generation costs for the PJM footprint. The study's findings are striking and deserve careful attention. The study estimates overnight[7] capital costs for the 2028/2029 Delivery Year at $1,361/kW for a combustion turbine (CT) and $1,419/kW for a combined cycle (CC) plant.[8] Translating these into the levelized first-year revenue requirement (Gross CONE), the study reports $663/MW-day for a CT and $813/MW-day for a CC – representing increases of approximately 47% and 44%, respectively, in real terms compared to the prior quadrennial review conducted just 2.5 years earlier in 2022. The drivers of this escalation include: higher equipment costs for the new GE 7HA.03 gas turbines (even net of some economies of scale), higher cost of capital reflecting rising interest rates since 2022, extended construction periods that delay revenue recovery, and the loss of bonus depreciation due to changes in tax law.

Importantly, the Brattle study explicitly acknowledges that these administrative estimates – while substantially higher than the 2022 study – still likely understate what developers will actually need to recover to justify new investment in the near term. The study presents three distinct cost concepts: (1) current-level nominal CONE incorporating today's premium pricing; (2) long-run equilibrium CONE based on 2022-era equipment costs with updated financing; and (3) short-term reservation prices that reflect what investors actually need to earn a positive net present value given that supply chains are tight today but expected to normalize over time. The short-term reservation prices – reflecting real developer decision-making – imply a CT price of $870–$2,779/MW-day and a CC price of $633–$2,070/MW-day depending on how long the developer expects current scarcity conditions to persist. While Brattle did not recommend the short-term reservation prices as a VRR Curve reference point, this range illustrates the degree to which administrative CONE estimates understate the actual market-clearing prices needed to attract investment under current conditions.

Empirical Costs: What Developers Are Actually Paying

The divergence between administrative and empirical cost estimates is not merely a technical footnote – it reflects a gap in what the market requires to trigger investment decisions. A September 2025 analysis by GridLab, *The New Reality of Power Generation: An Analysis of Increasing Gas Turbine Costs in the U.S.* (prepared with Energy Futures Group and Halcyon), assembled actual capital cost data from recent Certificate of Public Convenience and Necessity and Integrated Resource Planning filings across the United States – the most granular and current dataset of its kind, since the most detailed project cost data is rarely accessible.[9]

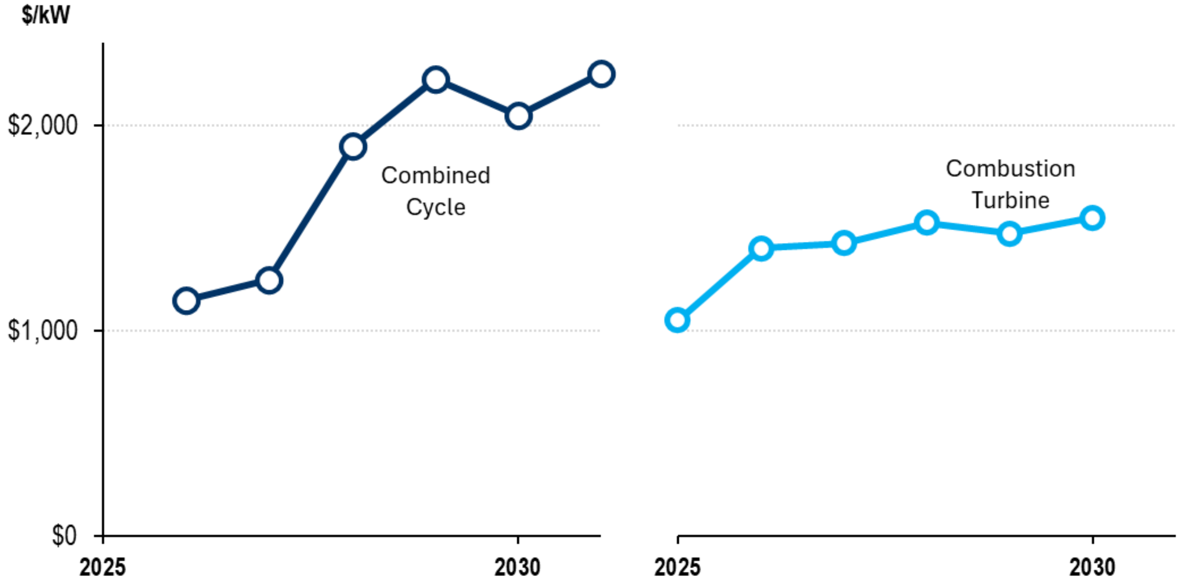
---

[7] The overnight cost is the theoretical cost of a resource if it could be constructed instantly ("overnight") incurring no interest during construction.

[8] CONE estimates here provided for CONE Area 3 (Rest of RTO). The Brattle Group and Sargent & Lundy, Brattle 2025 CONE Report for PJM (PDF) (April 2025), p. 8, Table ES-1.

[9] GridLab, Energy Futures Group, and Halcyon, *The New Reality of Power Generation: An Analysis of Increasing Gas Turbine Costs in the U.S.* (September 2025), pp. 1–2

Figure 2. Capital Cost Escalation of Natural Gas-Powered Resources[10]



The GridLab data tells a clear story. Combined cycle projects slated for completion in 2026–2027 were reported at capital costs of $1,100–$1,400/kW. By contrast, CC projects with 2028–2031 commercial operation dates are routinely reporting costs of $2,000/kW or more, with several PJM-specific projects in the data set in the range of $2,100–$2,300/kW.[11] A linear regression of the dataset confirms that costs increase with commercial operation date at a statistically significant rate – the later the project, the higher the cost – consistent with a market in which equipment and construction backlogs are compounding over time. Representative examples from PJM and adjacent markets include:

	Brown 12 – **PJM**	Homer City Generating Station – **PJM**	Cayuga CC Project – **MISO**	Chesterfield Energy Reliability Center – **PJM**
Type	CC	CC Redevelopment	CC	Simple cycle CT
Size	645 MW	4,500 MW	1,476 MW	944 MW
Commercial Date	2030	2028/2029	2029/2030	2029
Capital Cost	$2,144/kW	$2,222/kW	$2,256/kW	1,557/kW
Total Capital	**$1,383M**	**$10B**	**$3,330M**	**$1,470M**

The GridLab report further highlights that the public datasets most commonly used by regulators and market designers – including the NREL Annual Technology Baseline (ATB) and the EIA's Annual Energy Outlook capital cost assumptions – substantially understate current market costs. The 2024 NREL ATB projected a 2030 overnight cost of approximately $1,638/kW for a 1×1 combined cycle plant (moderate scenario, converted to nominal dollars); the EIA's Sargent & Lundy-based estimate for the AEO2025 placed a 1×1 CC at $921/kW in 2023 dollars ($1,058/kW

---

[10] Halcyon gas power plant tracker data as presented by contributor to GridLab (2025) report: Bullard, Nat. Decarbonization: Parameters, Dollars and Sense, Electrons Photons Molecules (January 2026), slide 87

[11] GridLab (2025), pp. 5–6 and Appendix Table 2

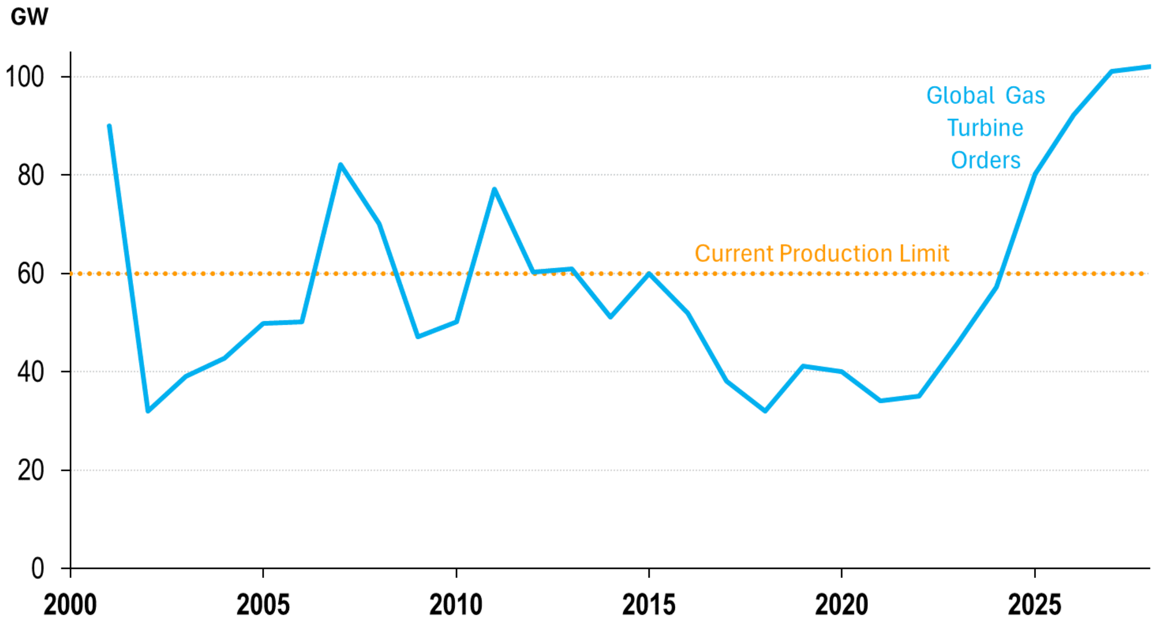
escalated to 2030 at 2% inflation). Both of these are materially below what developers are actually paying for comparable projects today.[12]

What Is Driving Cost Escalation – And Why It Will Persist

The cost escalation is not primarily a COVID-era supply chain disruption that will normalize quickly. Multiple converging structural forces are driving costs higher, and recent evidence from the major gas turbine OEMs suggests the trend is durable:

Gas turbine equipment scarcity. GE Vernova, Siemens Energy and Mitsubishi Heavy Industries are the three primary global suppliers of large industrial gas turbines. All three are reporting strong order backlogs and robust demand driven by global coal retirement, energy transition needs and the data center boom. Equipment scarcity has translated directly into OEM pricing power. The Brattle study found that gas turbine costs rose 15%–21% in real terms just between August 2024 and early 2025, on top of the 43%–46% real increase over the 2022 quadrennial review period as a whole.[13] Data on global gas turbine manufacturing capacity (Bloomberg, Barclays, Wood Mackenzie, IEA) shows that global orders are reaching approximately 100 GW per year by the late 2020s – two-thirds higher than current production limits of roughly 60 GW per year – suggesting that supply-demand imbalance in the equipment market may worsen before it improves.

Figure 3. Global Gas Turbine Manufacturing Capacity Compared to Current Demand[14]



Non-turbine components facing simultaneous cost pressure. The turbine is only one element of total plant cost. Transformers have seen sharp price increases driven by soaring costs for grain-oriented electrical steel and copper; switchgear prices have risen as reflected in the Producer Price Index for switchgear manufacturing; steel pipe costs

---

[12] GridLab (2025), p. 6, citing NREL (2024) ATB and EIA Capital Cost and Performance Characteristics Report (January 2024)

[13] The Brattle Group (2025), p. 3

[14] Bullard (2026), slide 28

more than doubled between 2020 and 2021 and have not fully retreated; and construction labor costs have increased due to skilled-worker shortages, prevailing wage pressures and inflationary dynamics.[15]

Extended project timelines compounding financing costs. Brattle explicitly models extended construction periods in its 2025 study, which increase the capital charge rate and therefore the Gross CONE even holding overnight capital costs constant. A combined cycle plant that requires 4–6 years from financial investment decision to commercial operation accumulates more financing cost during construction than a plant that could be built in three years. For projects in today's interconnection queue, timeline extensions due to study processes, equipment delivery and permitting approvals are themselves a source of cost escalation that does not show up in overnight capital estimates. The component-level data is stark: Lead times for Generator Step-Up transformers and gas turbines have each extended to three to four years, compared to the historical standard of roughly 18 months. The cumulative effect is that overall project development timelines have stretched from a historical baseline of approximately 24 months to four years even under extremely optimistic assumptions – a doubling that compounds carrying costs and exposes developers to additional market, regulatory and supply chain risk throughout the development cycle.[16]

Permitting as a binding late-stage constraint. Equipment and financing challenges interact with a third structural barrier: local permitting. PJM's tracking of milestone change causes for projects in active development (January 2023 through January 2026) found that permitting delays accounted for 29% of all reported milestone changes, compared to 23% for supply chain delays and 11% for transmission owner causes. Independent analysis of utility-scale wind and solar projects nationally estimates that at least 30% are canceled during the siting process due to local ordinances and community opposition. Unlike equipment lead times, permitting constraints cannot be addressed through reservation fees or accelerated procurement. A project that has cleared PJM's interconnection process and secured an executed agreement may still face years of regulatory proceedings at the local level – or outright denial – before a shovel enters the ground.

Financing costs and investor risk premiums. The Brattle study estimates an after-tax weighted-average cost of capital (ATWACC) of 9.5% for a merchant generation investment under current conditions – materially higher than in the 2022 study, primarily driven by higher interest rates. For a $1.5 billion combined cycle project, a 100-basis-point increase in the cost of capital translates to tens of millions of dollars per year in additional revenue required to service debt.

The Implication for Capacity Market Prices and Investment

The cost picture has a direct and important implication for capacity market pricing. The Brattle study recommends a VRR Curve reference price of $350/MW-day UCAP for the RTO (with higher values for EMAAC at $600/MW-day and ComEd at $725/MW-day), representing a midpoint estimate of long-run marginal cost between today's elevated level-nominal CONE and lower long-term equilibrium estimates. This recommended reference price is itself substantially higher than what the prior quadrennial review indicated – and yet it still likely understates what developers will require in the near term given the short-term reservation prices discussed above.

The practical consequence is that the capacity market, working as designed, will need to print prices at or above the administrative Net CONE for an extended period in order to attract entry of new supply. But as the paper's analysis in the following sections makes clear, those prices – while economically necessary – will collide with the political and affordability realities of a system where much of the native load appears not to have been hedged against capacity

---

[15] GridLab (2025), pp. 3–4, citing S&P Global Commodity Insights and U.S. Bureau of Labor Statistics Producer Price Index

[16] PJM Interconnection, 2027/2028 Base Residual Auction Reserve Target Shortfall Report (2025), pp. 20–23

price cycles of the magnitude now being observed. This is the essence of the credibility trap: The signal that is required to attract investment is the same signal that triggers the intervention that undermines it.

A final point deserves emphasis. The cost escalation documented here is not primarily the result of market design failure or regulatory dysfunction. It reflects genuine physical and economic scarcity: The global supply of gas turbines, specialized construction labor, electrical equipment and project development expertise is insufficient to meet the simultaneous demand from energy transition investment, data center buildout and generation replacement across the United States and globally. Even a perfectly designed capacity market cannot build a combined cycle plant in three years if the turbine manufacturer requires six. The market design challenge is not to lower costs through regulatory innovation; it is to structure incentives and risk allocation in a way that attracts the investment that is possible, as quickly as it is physically achievable, at a cost that ratepayers and policymakers can sustain over the decade-plus horizon that physical supply constraints will require.

Shared Reliability During Periods of Scarcity

For nearly a century, the shared reliability compact functioned as a form of mutual insurance – a community reservoir in which one utility or zone might lean on the system's shared resource adequacy reserves during a specific event, with the expectation that all participants contributed their fair share to the pool over the long term. Because load growth was historically slow, predictable and geographically uniform, this "leaky bucket" model worked. The risk pooling was viewed as equitable because the participants were largely homogeneous.

Today, a pronounced divergence is straining this pooling equilibrium. Load is entering the system at a pace that outstrips both the region's historical averages and the rate at which new supply can be developed and interconnected.

In a tight market, this divergence creates what economists call a "reliability externality." When a large new load connects to the system without commensurate new supply, it draws on the shared reliability margins that were sized and paid for under a different set of assumptions – effectively borrowing reliability from a pool it has not yet had the opportunity to replenish. To be precise, this is not a claim that new loads pay nothing: All loads pay capacity market charges. The externality arises from timing and pace. Capacity charges reflect costs as priced in the auction; they do not fully compensate existing customers for the reliability degradation that occurs when large new loads simultaneously arrive before the generation fleet has expanded to serve them.

Within limited bounds, this pooling dynamic is acceptable and even intended – the sloped VRR Demand Curve was designed to allow system reliability to vary modestly from year to year as market conditions shift. The externality becomes a genuine design problem when new load connects at a pace and scale that persistently outstrips the ability to bring commensurate generation online, driving a sustained shortage that degrades reliability for the entire footprint.

This strain raises a foundational allocation question: When the centralized market falls short, how should the available capacity be distributed?

- **Historical homogeneity:** Continue under the long-standing paradigm that treats all loads uniformly across zones, allocating procured capacity in proportion to peak load share and leaving further allocation choices to EDCs and states.
- **Deliberate differentiation:** Acknowledge that some principle – cost causation, contribution to supply or explicit contractual commitment – should govern which loads receive priority when capacity is genuinely scarce.

Pricing Capacity as a Scarce Good

The clearing prices in recent Base Residual Auctions have been at the center of the controversy over PJM's capacity market performance. Assessing whether those prices reflect market dysfunction or sound design requires first accounting for the underlying physical conditions: The system is in a period of structural scarcity driven by rapid demand growth, accelerated thermal retirements and supply chain constraints on new generation.

In a market with abundant supply, clearing prices track the cost of the marginal resource – the last unit of supply needed to meet demand. But RPM was intentionally designed to reflect the demand side as well. When physical supply falls short of the reliability requirement, prices no longer track the marginal cost of supply; instead, the Variable Resource Requirement (VRR) Curve – which represents society's declining marginal willingness to pay for reliability above and below the target reserve margin – governs where prices clear. This allows prices to rise above Net CONE during a shortage, signaling the social cost of unmet reliability rather than merely the cost of the next marginal unit of supply that may fall short of the empirical Net CONE.

It is sometimes argued that clearing prices should be administratively capped at Net CONE because new supply cannot immediately respond to the price signal. But the design explicitly rejected this approach. First and foremost, the design allows the price to clear below Net CONE in periods of surplus. Therefore, the price must also be allowed to rise above Net CONE in periods of scarcity to mathematically enable the long-term average revenue requirement equal to Net CONE needed for investment. Furthermore, above-CONE prices perform three distinct functions during a deficit that a binding cap would suppress:

1. **Retention:** In a shortage, the most immediate lever to maintain reliability is preventing further exits. The design relies on high scarcity prices to signal to high-cost resources (perhaps older resources or those facing high opportunity costs given opportunities to export capacity off-system) that their capacity is critically needed.

2. **Capital Acceleration:** While greenfield generation lead times are long, they are also often a function of capital intensity. The market was designed to allow high prices to incentivize developers to pay premiums for expedited equipment delivery, overtime labor and parallel permitting tracks, thereby shortening the time required to resolve the shortage. The effectiveness of this mechanism is constrained when shortages are driven in part by global equipment scarcity and regulatory permitting timelines rather than financing gaps alone – high prices can accelerate capital formation, but they cannot manufacture turbines or shorten interconnection queues.

3. **Unlocking Alternative Supply and Demand Response:** Allowing prices to ascend the VRR Curve is designed to test the elasticity of the market. It activates demand response (i.e., customers self-identifying a willingness to reduce their consumption of reliability in exchange for cost reductions). Further, a high cap prevents "artificial scarcity," ensuring that high-cost existing or new generation (whose legitimate, mitigated avoidable costs might sit above a suppressed cap) can clear the market and be retained for reliability.

Because physical supply fell short of the reliability requirement in recent auctions, clearing prices naturally rose along the VRR Curve. The price collar added a secondary constraint: Over 800 MW of capacity that provided cost-based offers above the $333.44/MW-day administrative cap could not clear – a direct illustration of how price administration can exclude cost-reflective supply even when physical capacity exists and is willing to serve.[17] In this context, the

---

[17] PJM Interconnection, 2027/2028 Base Residual Auction Reserve Target Shortfall Report (2025), p. 5

capacity market is doing what it was designed to do: communicating the high cost of reliability during a period of structural scarcity. That signal is economically necessary. But as the following section explores, its social and political consequences create their own structural problem.

The Gap Between Price Signals and Durability With Government

Scarcity pricing is economically necessary, but its effectiveness as an investment signal depends on durability – on investors' confidence that high prices will persist long enough to attract and retain the capital they are meant to signal. That durability is precisely what the current hedging environment puts at risk.

A substantial share of PJM load is served through short-term default service procurements rather than long-term bilateral hedging. When capacity prices spike to signal scarcity, that volatility passes through to unhedged consumers with little lag, creating an affordability shock that generates intense pressure for regulatory intervention. This dynamic creates what this paper calls the *credibility trap* – a feedback loop that undermines the very investment the high prices are meant to solicit:

The Signal: Prices rise to signal scarcity (e.g., reaching the price cap).

The Reaction: Because a significant share of PJM load is unhedged, the price spike triggers immediate consumer pain and regulatory scrutiny.

The Risk Premium: Investors, observing the backlash – including actual regulatory interventions, backstop procurement mandates and price cap proposals that have emerged in the PJM region – discount the durability of the revenue stream. Even where intervention is contemplated but not finalized, the uncertainty itself dampens investment. LSEs simultaneously become disincentivized to enter long-term contracts that could support new entry, uncertain whether today's high prices will persist or be muted before those contracts can recover their costs.

The Stagnation: The market signals "build," but the political environment signals "risk." Capital stays on the sidelines or moves to other markets, and the shortage persists.

This is the structural paradox at the core of PJM's current situation: Prices must be high enough to incentivize supply in a constrained environment, but the state of load hedging makes those prices unsustainable long enough to do their job. The market is generating the price signal it was designed to generate during a shortage. The gap between design intent and observed reality is not in the pricing mechanism – it is in the assumption, embedded in the original design, that most load would be insulated from spot volatility through long-term forward contracts.[18] Governmental

---

[18] PJM's visibility into the full extent of load hedging – including hedging embedded in default service procurement bids – is incomplete. The characterization of "unhedged load" describes revealed affordability exposure and observed political dynamics, not a precise count of unhedged megawatts.

pressure on electricity prices is not new – it has characterized regulated energy markets for a century – but the combination of high prices, low hedging depth and an accelerated cost environment has compressed the timeline for the feedback loop to become damaging.

The Limits of Economic Rationing: The Social Priority Paradox

In a textbook commodity market, scarcity is resolved through pure economic rationing. As supply tightens, the price rises until those with the lowest “willingness to pay” voluntarily exit the market. The commodity flows exclusively to the buyer who values it most highly in financial terms.

However, applying pure economic rationing to physical electricity delivery would produce outcomes that are both socially and politically untenable. Electricity serves hospitals, emergency services and residential customers whose reliability requirements are defined by public necessity rather than market participation. A framework that results in a neighborhood going dark during a shortage because an industrial facility had outbid it for a higher reliability tier would face fundamental legitimacy problems regardless of its economic efficiency.

To prevent this, the region has long relied on the shared reliability compact and a consistent standard of reliability across the zones within PJM’s footprint. This is inherent in the RPM design today, where the pool-wide centralized procurement does not allocate the procured capacity or physical reliability only to the highest bidder, but instead provides a uniform standard of reliability for all zonal loads.

Notwithstanding, it still sought to apply fundamental market principles to the pricing of it. The design assumed that prices would be allowed to rise significantly above the long-run equilibrium cost of new entry during scarce conditions. The implicit assumption was that native load would be protected from these cost impacts not through suppressed price caps, but through a combination of long-term hedging and voluntary demand response (which allows consumers to self-identify their low willingness to pay by opting to curtail load in exchange for compensation).

The rapid growth of hyperscale data centers is straining this compromise in a way its designers did not anticipate. The compact’s design assumed that new loads would enter the system gradually and in rough proportion to the growth of supply. Today, large loads are interconnecting at a pace that outstrips available supply, pushing clearing prices up the VRR Curve. Simultaneously, the long-term hedging arrangements that were supposed to insulate native load from the resulting price volatility have not been adopted consistently across the load base.

The consequence is a sustained affordability shock that has generated governmental pressure of a different character than occasional scarcity spikes: pressure to revisit the market’s basic architecture rather than merely adjust its parameters. That pressure reflects a question that was never considered when the original compact was made: whether loads that impose a net strain on the system’s reliability margins – even while paying the going market rate – should have the same standing in the shared pool as loads whose presence preceded the scarcity they now share.

PJM’s centralized market was designed on the assumption that cost and reliability would be shared broadly and that the compact would be self-reinforcing. Today that assumption is being tested from two directions simultaneously: Load growth is outpacing supply, and the hedging arrangements that were supposed to absorb the public backlash to cost volatility are inconsistently in place. The current framework is straining because it is being asked to manage a tension it was not designed to resolve: how to simultaneously serve market participants with fundamentally different reliability needs, cost structures and political standing on a system that cannot physically discriminate between them in real time.

This tension frames the three paths explored in the sections that follow. The central question each path must answer is whether the shared reliability compact – the promise of a common grid reliability standard for all customers – can be preserved, and if so, at what cost, through what institutional mechanism and for which customers?

IV. Three Paths Forward: Choices the Region Can No Longer Defer

▸ **Key Takeaways From This Chapter:**

- Three high-level paths forward are presented for discussion to help confront the resource adequacy and investment challenges facing the region today.
- The first approach would stabilize both customer bills and investment signals by having most electricity supply be contracted through longer-term agreements – a model with established precedent in regulated utility frameworks and bilateral contract markets – though it raises important questions about how risk is shared between consumers and investors, and how much procurement flexibility states and suppliers retain.
- The second approach would make explicit what today happens by default – that when supply falls short, some customers may bear more of the burden than others – and would establish principled rules for that allocation based on who caused the scarcity or other policy objectives; near-term application focuses on new large loads that connected without contributing new supply, though the broader question of whether states, customer classes or regions should have differentiated reliability is one that deserves careful deliberation across all affected parties.
- The third direction would gradually shift the mechanism by which generators recover their costs away from the administered capacity market and toward the energy market itself, paired with requirements that customers must be hedged through long-term energy contracts – an approach that may be more structurally durable, but one that requires sustained commitment across regulatory jurisdictions, careful sequencing, and a transition timeline measured in years rather than months.

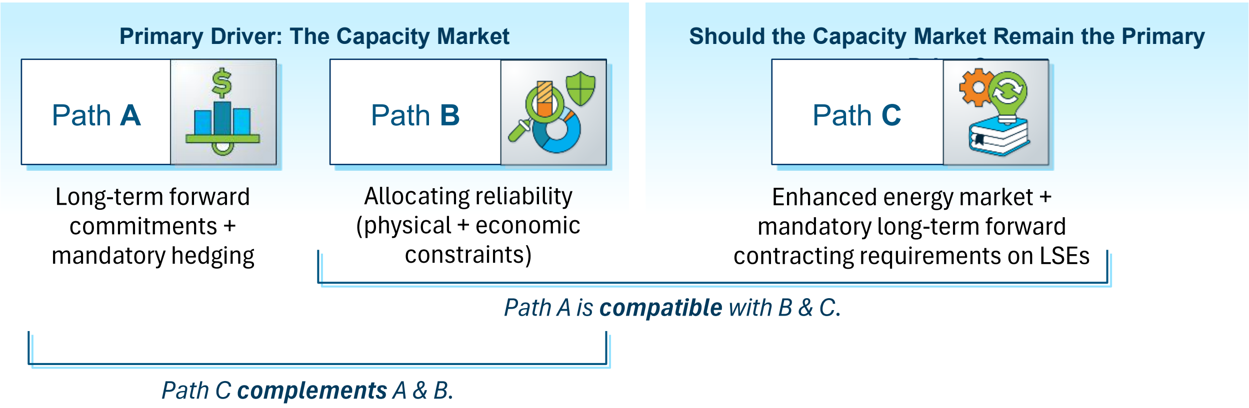
Additional Reforms That Advance the Region Under Any Path:

- Regardless of which path the region chooses, several reforms are overdue and should be explored in parallel: capacity market reforms including seasonal procurement designs and accreditation refinements, and a package of energy and reserve market reforms to accurately value operational flexibility and send price signals that reflect the true cost of reliability – none of these wait on the longer-horizon path decisions.
- The large loads straining the system today also represent an opportunity: Hyperscale data centers have genuine flexibility – AI workloads can be shifted; on-site storage or generation can be deployed. That flexibility could be accessible economically when the price is high enough to make it worthwhile; getting the price signals right unlocks a demand-side resource that could meaningfully ease the transition.

The current disequilibrium in the PJM footprint forces us to confront uncomfortable trade-offs. There are no simple solutions to a capacity shortage. However, this crisis offers a rare opportunity to reexamine the foundational assumptions and objectives of the region's resource adequacy mechanism.

For two decades, we have operated under the assumption that the policy objectives valid at the inception of the Reliability Pricing Mechanism (RPM) – namely, coordinating entry/exit decisions around the margin, facilitating retail choice through short-term signals and, at the highest level, reliability at least cost – would remain valid indefinitely. Today, we must test those axioms. Rather than assuming the status quo is the only baseline, this paper explores alternative combinations of internally consistent assumptions.

Figure 4. Three Paths Forward



Each of the following proposed frameworks are conveyed in a manner to explain how they may work at a high level and how they seek to address some of the challenges being confronted today. Underlying each is a significant amount of detail that requires careful development and consideration. PJM presents these to begin the discussion.

Path A: Stabilized Markets (The "Come Hedged" Model)

This path preserves the long-standing compact of shared reliability in PJM by making it financially durable.

Under this path, the capacity market continues, but the vast majority of load is required to be covered through long-term forward commitments either through mandatory LSE hedging requirements or through centralized, long-term procurement of capacity. In either case, the purpose would be to procure the vast majority of capacity needed to maintain resource adequacy prior to the final auction for the delivery period, or capacity spot market, so that it may print high scarcity prices when the system is short (maintaining the investment signal), but sufficiently insulates most load from those prices through forward contracts or commitments. This path trades the optionality of short-term procurement for the stability needed to support investment, and in doing so, seeks to avoid the credibility trap.

Durability Through Hedging

In this framework, the capacity spot market continues to serve its vital economic function: sending price signals consistent with supply and demand fundamentals in PJM including printing high prices to signal scarcity. However, the impact of those signals is fundamentally altered. Under this path, the vast majority of load would be protected from spot volatility through long-term contracts, either arranged bilaterally by the LSE or procured on their behalf via an altered capacity auction structure.

The objective here is nuanced. We would not aim to artificially lower the cost of capacity.

In a scarcity environment, equilibrium prices must remain high (above Net CONE) to attract new entry and retain existing resources that might otherwise retire or export. Instead of avoiding this outcome, we would aim to eliminate the shock of that cost by shifting the financial obligation from a potentially volatile annual auction to a portfolio of long-term contracts that smooth the ratepayer impact over a long period of time.

This distinction is critical. To attract new entry in the near term, the market must support prices high enough to compete with the opportunity costs of data center collocation or export to neighboring systems. However, while a price of $300–500/MW-day or more might trigger a public political crisis if it hits ratepayers unexpectedly in a single delivery year, that same cost becomes more manageable if it is approached gradually. This path essentially asks consumers to trade the potential upside of low spot prices for the benefit of being protected against high (possibly sustained) spot prices and resulting affordability concerns, and improved certainty of long-term reliability.

One potential outcome of this mechanism is that the system avoids sustained periods of shortage because the forward hedging required of the load stimulates investment in physical resources to satisfy those hedges. However, if this is not the case, the only load exposed to those high prices is the relatively small amount of unhedged load, creating an incentive for that load to more completely hedge itself in the future in order to avoid exposure to those high prices. The most natural provider of that financial hedge is a resource owner, whose physical capital asset provides them with a physical hedge to their financial position. This incentive helps drive investment in new resources, which causes prices to come back down. If prices are not allowed to rise high enough to drive this behavior, then the enforcement challenge described below arises.

The Enforcement Challenge

A “forward showing” requirement is only as strong as the penalty for noncompliance. If an LSE fails to procure its required capacity three years out, what is the consequence?

If the consequence is solely financial (i.e., forcing the LSE to pay the spot clearing price) we return to the circular arguments of the current construct. If the spot price (or penalty) is capped to protect consumers (the VRR cap), then the consequence of failing to pre-arrange supply is insufficient to deter the behavior. An LSE might rationally choose to remain short, pay the capped penalty and lean on the system’s reliability. Therefore, a true stabilized markets approach requires a compliance penalty that effectively acts as a super-cap – a price high enough to incentivize sufficient long-term contracting and hedging to be the only rational business decision. We cannot escape the necessity of valuing reliability at a premium in this framework, but we can change when that premium is paid.

A second potential consequence is levying the physical penalty. An LSE that is short would either have its load shed first in an emergency or receive a larger portion of the overall system load shed. There are technical limitations to the ability to implement this as explained previously; however, there is some level of granularity, likely at the level of the transmission owner, where targeted load curtailment could be implemented. This is discussed further under Path B.

A meaningful transition period will also be necessary, and its design matters as much as its destination. Directing large volumes of previously unhedged load to procure bilateral capacity contracts simultaneously in a market that is capacity-short and more concentrated in ownership than in prior cycles creates its own risk. A hedging mandate phased in too quickly, without commensurate attention to market power mitigation, could exacerbate the affordability harm it is designed to prevent. The IMM’s oversight role would be especially important during any such transition.

The Role of PJM and Evidence of Feasibility

In this framework, PJM’s role could evolve from auctioneer to a facilitator of durability. This may involve standardizing long-term reliability products or, in states that lack the mechanism to sign ten-year power purchase agreements, administering a centralized procurement of long-term capacity contracts on their behalf.

While this shift requires a significant departure from the relative “short-termism” of the current RPM, there is ample evidence that high capacity costs are politically sustainable when they are properly hedged.

California ISO (CAISO)

The California resource adequacy market has seen weighted-average prices for capacity rise significantly in recent years, often exceeding PJM's record highs. Yet, because these costs are absorbed through bilateral contracts and regulatory proceedings rather than a single transparent spot auction, the market continues to function without the collapse of the regulatory compact.[19]

Vertically Integrated States

In certain regulated jurisdictions (even in PJM), the "cost" of capacity – including the recovery of investments in existing resources to ensure reliability – is often substantial. However, because these costs are rate-based and amortized over years, they are accepted as the necessary price of keeping the lights on. For example, the regulated capacity rate for load served under the FRR by Appalachian Power Company in 2020/2021 was $480.98/MW-day,[20] while the RPM RTO capacity price in the BRA for that delivery year was $76.53/MW-day.

One lesson these examples suggest: The problem is not simply the high cost of reliability; it is the volatility and exposed risk of the unhedged spot signal. Path A proposes that we solve this by making the "hedge" mandatory.

Improving Investability

The cost escalation documented in the preceding section is one dimension of the investment challenge. The other is structural: Even if prices are high enough to theoretically justify investment, the economics of project financing require that those prices be *credible* over the financing horizon. This is where the current market design creates friction.

Long-term revenue certainty is a structural requirement, not a preference. When a generation developer finances a $2 billion combined cycle plant with project debt, lenders underwrite the transaction based on projected revenue streams over a 15- to 20-year debt tenor. Capacity market revenues that are determined in annual or three-year forward auctions under rules that have changed materially multiple times in the past decade – MOPR expanded and reversed, ELCC methodology revised, auction schedules delayed – cannot be treated as durable over that horizon. The response from capital markets has been predictable and rational: Project finance now requires long-term off-take agreements, power purchase agreements or other contractual revenue certainty before underwriting at manageable debt costs. Without such agreements, developers must either accept higher equity returns (raising the effective cost of generation), rely on balance-sheet financing (limiting the pool of potential developers) or walk away in favor of opportunities with clearer revenue visibility.

Regulatory stability compounds the certainty problem. Higher capacity prices are necessary but not sufficient. The same administrative changes that have repeatedly altered capacity market rules over the past decade also signal to investors that future changes are possible – that the revenue stream they are counting on when they make a financial investment decision may be modified by the time the plant generates power. This "stroke-of-the-pen" risk extends beyond market rules to the broader state and federal policy environment. State-level clean energy legislation, federal EPA rulemakings, and shifts in permitting policy for offshore wind and pipelines each represent vectors through which the economic viability of a specific resource type can change materially within the debt tenor of

---

[19] California's CPUC exercises active oversight of long-term bilateral procurement terms and conditions – this regulatory involvement is itself part of why high costs are absorbed without political disruption and a design feature worth considering in any centralized PJM procurement model.

[20] AEP Appalachian Power Company capacity formula rate summary for the 2020/2021 Delivery Year (PDF)

a project finance structure. Investors discount all of these vectors into their required returns, raising the effective cost of capital even when spot auction prices appear adequate.

The implication for market design is direct: The structural economics of generation investment require that market design reforms address not just the *level* of revenue but the *durability* and *contract form* of that revenue. High spot auction prices that remain exposed to policy uncertainty do not solve the investment problem. Long-term commitments – whether through bilateral contracts, mandatory forward hedging requirements or PJM-administered long-term procurement – are the mechanism through which investment economics and market design can be realigned.

This analysis is corroborated by PJM's direct engagement with various entities in the generation development life cycle in and outside of the PJM footprint. In conversations with generation developers, project finance lenders, equity investors and power project consultants in furtherance of this initiative, PJM heard a strikingly consistent message: High capacity prices alone do not drive investment decisions. The barriers cited – repeatedly and across investor types – were revenue durability over the project financing horizon and regulatory stability sufficient to make long-term revenue projections credible. As an example, one entity observed that in the time between the decision to build a new fossil resource and when it generates its first megawatts, the state and federal governmental officials setting policy that affect the operation of the resource will change at least once and maybe twice. Given the volatility of those policies, it is incredibly risky to extend the significant capital needed to build a new resource without some longer-term agreement that creates revenue certainty. This consistency across otherwise competing interests reinforces that the barriers described above are structural features of the current investment environment, not idiosyncratic concerns of particular participants.

The following considers potential market constructs under Path A for consideration, although certainly not intended to be an exhaustive representation of designs.

Option 1: LSE Hedging Requirements

This option would help ensure the BRA is truly a "residual" procurement and limit LSEs' exposure to the auction spot price by requiring all LSEs come to the BRA with a minimum level of their expected forward capacity obligation hedged (e.g., 90%). The level and forward period of the required hedge would be up for debate but would likely need to cover a significant portion of the LSEs' forecasted position to adequately address the durability concerns of today.

To provide this information, LSEs would submit a resource plan to PJM ahead of the BRA with their owned or bilaterally contracted capacity for the future delivery year. This would be similar to resource plans required by FRR entities today, but only for the required hedging percentage of the LSEs' expected obligation.

The penalty for LSEs that failed to adequately hedge their forecasted capacity obligations and meet the required minimum would need to be substantial to help ensure LSEs satisfied their hedging requirements – at least as high as the price cap used in the spot auction. In turn, the critical design element under Path A (in having LSEs sufficiently hedged going into the spot market) would be that the demand curve price would be allowed to rise sufficiently high that it adequately deters LSEs from overly leaning on the rest of the system and sacrificing pool-wide reliability during periods of scarcity, without creating the level of affordability concerns and public backlash observed today. More succinctly, it needs to be more painful to buy from the market when short than it is for the LSE to cover their position physically.

Effectively, in order for all LSEs participating under the centralized procurement of RPM to view the clearing outcomes as just and reasonable, the spot price must be capable of rising high enough such that the collective maximum willingness to pay at the demand curve price cap is consistent with the LSEs' willingness to accept a lower reliability outcome (i.e., reach a sufficiently high price that the load would willingly be demand response and subject to more frequent curtailment than pay a higher price for incremental capacity and reliability).

The hedging requirement on LSEs would help address the “affordability” problem of today by limiting their customers’ exposure to the volatility and risk of sudden price spikes under single-year pricing, as well as help provide long-term price certainty for investment to support reliability.

Option 2: Mandatory & Centralized Long-Term Procurement by PJM on Behalf of LSEs

In this approach, PJM would expand the scope of the current RPM auctions and conduct a centralized procurement for a portion of the capacity needs on a longer-term basis (longer procurement horizon and period than the BRA).

As one example, PJM could procure 10% of capacity needs over 7-year contracts (i.e., for delivery years 3 to 9 in the future under the current three-year forward BRA design). Over time, this process would provide about 70% of all capacity obligations, with the remaining procured through the annual terms of the current BRA design.

This approach, like LSE hedging requirements, would help provide the needed price certainty to support investment and long-term reliability, while having LSEs and their customers less exposed to price volatility due to the laddering of the long-term contracts over time.

However, there are many critical design elements that would need to be answered for PJM to implement this structure around long-term contract terms, including:

1. The level of capacity to procure under the long-term contracts (i.e., more or less than the 70%)
2. The length of the contract terms (i.e., more or less than the 7 years)
3. Willingness to pay for the capacity, among several other important elements

Relative to LSE hedging requirements, this approach provides less flexibility for LSEs and market participants in setting their own long-term contract terms but may work better for restructured states that currently have less visibility into the forward obligations of individual LSEs under retail choice.

International Support for Long-Term Contracting

Globally, the transition toward long-term capacity contracting reflects a growing consensus that short-term price signals are often insufficient to de-risk the massive capital required for new build. In Great Britain and Italy, centralized capacity auctions offer 15-year “new build” contracts that provide the revenue stability necessary to secure project financing for both dispatchable and low-carbon resources. France has recently transitioned to a centralized capacity auction and uses long-term tenders for new-build, and Japan has pushed the boundary with its Long-Term Decarbonization Auction (LTDA), which utilizes 20-year contracts for capacity to incentivize emerging technologies like hydrogen-ready gas and long-duration storage. Similarly, Germany is pursuing both long-term contracting for new-build to support its 2026 tenders for hydrogen-ready capacity and a transition to a centralized capacity market that has proposed long-term contracting for new-build. These international frameworks demonstrate a strategic shift away from pure merchant risk, favoring “bankable” reliability where long-term regulatory commitments serve as a necessary bridge and provide the required certainty for investment.

Trade-Offs To Consider

It is important to recognize that while the above options can help provide longer-term price certainty to support resource investment and reliability, as well as help address today's affordability, they do come with trade-offs.

First, the RPM was intentionally designed to limit the centralized procurement to a single delivery year in the future with the assumption that longer-term bilateral contracting between willing buyers and sellers would naturally be encouraged to hedge forward positions, under mutually agreeable terms and conditions, and without an assumption that one size fits all. Although this wasn't fully realized, the value of that design still exists, as it enables parties to determine the level of hedging that best fits their company's risk profile and forward expectations.

As such, the market design would ideally keep the level of longer-term forward hedging voluntary. Unfortunately, the recent history has shown that adequate long-term contracting and hedging has not occurred in the market when allowed to be voluntary, as evidenced by the level of customers exposed to the recent hike in RPM Auction prices and affordability concerns that drove governmental intervention.

Second, a mandatory longer-term procurement of capacity or required hedge impacts risk.

On the supply side, longer-term commitments may increase deficiency risk: the longer the term of a resource's commitment, the greater the chance that an aging resource fails irrecoverably during the term of the commitment, leaving the resource owner and the system short. On the demand side, customers face greater risk of needing to continue supporting resources that become uneconomic during their commitment period.

More broadly, longer commitment periods shift technology, economic and policy risk from investors – who are best positioned to price and manage those risks – toward consumers. This is not an argument against longer-term procurement, but a design constraint that contract terms and risk-allocation mechanisms must address explicitly: Who bears the cost if a resource's economics deteriorate substantially during a multiyear commitment? Conversely, longer-term commitments also help de-risk project investment and can enable new generation to enter the market at lower prices providing cost savings to consumers.

An additional consideration in deciding if the wholesale market design should move in this direction is if there have historically been structural or regulatory barriers in place that drove the lack of efficient long-term hedging for certain LSEs, and in particular, those in restructured states. A significant portion of retail load in restructured states is supplied under default service procurements arranged by the EDCs and regulated by the utility commissions. The contracts for the default service typically have durations of three years or less, choosing to rely on relatively short- or intermediate-term contracts.[21] Additionally, for some states, the ability to enter longer-term contracts has been limited by statute.

If this has been a primary barrier to longer-term contracting for certain LSEs, and the retail customers under default service represent those that were most impacted by the sudden price surge and affordability concerns, the most direct solution may be to adjust the default service procurements in a manner that better promotes longer-term contracting and price stability for customers. Alternatively, this could be a necessary change in addition to the wholesale market reforms discussed under Path A.

---

[21] These shorter terms were not an oversight; they were deliberately designed to support retail choice by preserving customer optionality and avoiding lock-in through the default service. Extending default service procurement horizons would require explicitly weighing this trade-off: Longer terms improve affordability, stability and investability signals but may inhibit the customer switching that retail choice programs were designed to promote.

Lastly, the following options would continue to be available under Path A to help LSEs and customers manage their capacity obligations and costs:

- **Demand Response:** Since its inception, the amount of voluntary demand response has grown substantially, recently providing around 8 GW of capacity value to the system. The demand response product provides consumers with the ability to reflect their own willingness to pay for capacity and the associated reliability. If the price of capacity becomes too high for the value provided to a customer or aggregation of customers, they can work with a curtailment service provider to participate in demand response to express their individual willingness to pay for capacity at a price below the collective maximum price represented on the VRR Curve. This option would continue to exist under the mandatory hedging options discussed under Path A, as well as would continue to provide an avenue for LSEs and consumers to manage their capacity obligations and prices in the spot market for the remainder of their unhedged loads.
- **FRR Alternative:** The FRR Alternative would still be an available option for eligible LSEs that prefer to fully manage their own resource adequacy obligations of the PJM pool and capacity costs rather than participating in the centralized procurement under the RPM Auctions. This option is effectively a complete hedging, where the FRR Entity is responsible for meeting their entire capacity requirement through self-supply (owned or bilaterally contracted capacity) and submitting their resource plan to PJM.

Simple Comparative Example

The following example is intended to illustrate how the capacity costs flow across a range of years for an LSE that enters into long-term hedging arrangements to cover a significant portion of their load (roughly 70% in this example) vs. an LSE that provides no long-term hedging for its customers.

LSE A:		
LSE A is assumed to cover about 70% of their forward capacity obligation through long-term contracts.	These contracts reflect an estimated long-term price of capacity at $250/MW-day over a 10-year period.	Assuming a 100 MW capacity obligation for LSE A, this example has 70 MW covered in long-term contracts at $250/MW-day, resulting in a daily cost of $17,500.

LSE B	
LSE B is assumed to enter the BRA with no forward hedging or long-term contracts.	LSE B also has a 100 MW capacity obligation and pays the BRA spot price for its full obligation each year.

Table 1. Illustrative Capacity Costs ($/MW-day) by Year for LSE A and B

Year:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Avg.**
BRA Price	$220	$180	$160	$100	$60	$30	$300	$450	$500	$500	$250
Load Obligation	100	100	100	100	100	100	100	100	100	100	100
BRA Costs	$22,000	$18,000	$16,000	$10,000	$6,000	$3,000	$30,000	$45,000	$50,000	$50,000	$25,000
LSE A:											
Contracted MW	70	70	70	70	70	70	70	70	70	70	70
Contracted Costs	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500	$17,500
BRA Revenues	$15,400	$12,600	$11,200	$7,000	$4,200	$2,100	$21,000	$31,500	$35,000	$35,000	$17,500
Total Net Costs	$24,100	$22,900	$22,300	$20,500	$19,300	$18,400	$26,500	$31,000	$32,500	$32,500	$25,000
LSE B:											
Total Costs	$22,000	$18,000	$16,000	$10,000	$6,000	$3,000	$30,000	$45,000	$50,000	$50,000	$25,000

	Avg. Costs	**Min Daily**	**Max Daily**
LSE A	$25,000	$18,400	$32,500
LSE B	$25,000	$3,000	$50,000

Impact: LSE A forward hedging significantly reduces exposure to the BRA spot price and volatility.

In this example, both LSEs pay the same total net capacity charges over the 10 years that average to $25,000 per day. However, the minimum and maximum costs vary significantly between them, with LSE A paying more in years where capacity is in surplus and receiving significant savings in years of scarcity compared to LSE B.

This is one example of many across which results will vary, but it is one that illustrates the value of long-term hedging for LSE A by avoiding sudden and significant swings in capacity costs relative to LSE B and reducing their customer's exposure to high scarcity prices during periods of shortage.

Path B: Differential Reliability (Rationing of Capacity in Scarcity Conditions)

Path B begins with an observation that is uncomfortable but worth stating plainly: reliability differentiation is not a new idea that requires invention. It already happens. Path B assumes that the investment needed to return the PJM system to its desired reliability level will either not happen in a timely manner and therefore it becomes a bridge solution, or, not happen at all due to affordability concerns, environmental policies or other such barriers that are valued to a higher degree than the 1-in-10 standard, for example.

When the system is short and load must be shed, someone gets cut before someone else. Today, that ordering is determined by distribution system architecture, operational conventions and emergency protocols that have little to do with which customers value reliability most, which loads imposed the greatest cost on the system or what any stakeholder has explicitly chosen. The question Path B raises is not *whether* to differentiate reliability – it is whether the region should make that differentiation *deliberate*, *principled* and *governable*, rather than leaving it to the physics of the power system and emergency operator discretion.

This reframing matters. If the "shared reliability" compact – the promise that every customer in the PJM footprint receives meaningfully equivalent grid reliability – cannot be maintained as the gap between projected load growth and feasible new supply widens, then the region faces a choice between two unsatisfying alternatives:

- **Implicit universal degradation:** The reliability standard erodes as reserve margins tighten. The "1-in-10" aspiration becomes a "1-in-3" or "1-in-2" reality for everyone, without any explicit policy decision having been made. This is the path of least political resistance and the greatest practical risk.
- **Explicit differentiation:** The region acknowledges that scarcity must be rationed and chooses to make that rationing transparent, consistent with cost-causation principles, and accountable to the stakeholders who bear its consequences.

Path B is the second of these. It is a response to a physical and economic constraint that the other paths may not fully resolve in time. And critically, it requires confronting a governance question that is at least as hard as the technical one: *Who has the standing to decide how reliability is allocated when it cannot be universalized? On what basis should that decision be made? Through what institutions and processes?*

The Dimension Already Being Worked: New Load Without New Supply

PJM is currently engaged with stakeholders in the Connect and Manage Senior Task Force on a framework that represents one principled answer to this question. Similar frameworks have been discussed under different naming conventions in other regions of the United States, including at least in ERCOT and SPP.[22] The core logic is straightforward: Large new loads that seek to interconnect to the system without bringing commensurate new generation capacity are, in a meaningful sense, free-riding on the reliability provided by the existing fleet. If the system is short and load must be shed, there is a defensible argument that these loads – which imposed incremental demand on a strained system without yet being able to contribute commensurate supply – should be curtailed before legacy load that has been served by and paid into the shared reliability pool.

A Connect and Manage approach operationalizes this principle by establishing curtailment priority rules that place new large load interconnections without associated new supply at the front of the curtailment queue. This is not a

---

[22] See Texas Senate Bill 6, 2025.
Also see proposed "Conditional High Impact Large Load" service in Southwest Power Pool, Inc., Tariff Filing, Docket No. ER26-1323-000 (filed Feb. 10, 2026).

radical invention: Cost-causation-based service differentiation already exists within states and at the distribution level, where utilities have long offered interruptible tariffs, curtailable service contracts and demand response programs. The novelty at the transmission level is the scale of the loads involved and the operational capability to curtail them. Hyperscale data centers and other large transmission-connected loads that make up much of the new load queue are often discrete enough to be curtailed directly by PJM operations without cascading effects on the surrounding distribution system – a physical controllability that simply did not exist when RPM was designed.

This approach addresses the most acute dimension of the reliability externality. It prevents new, unbacked load from diluting the reliability of legacy customers who funded and depend on the shared pool. It also creates a meaningful market signal: Large new loads that want firm reliability standing must bring, finance or contract for new generation capacity to support their interconnection. That alignment of incentives with cost causation is a genuine improvement on the current undifferentiated system.

The Harder Questions: Beyond New vs. Native

The Connect and Manage framework addresses one axis of differentiation – new incremental load versus legacy native load – through a cost-causation principle that most stakeholders can recognize as legitimate. But it does not resolve all of the questions that a more comprehensive Path B regime would eventually require the region to answer. The further one moves along this path, the harder the questions become – and the more they implicate institutions beyond PJM.

Geographic differentiation. The PJM footprint spans thirteen states and the District of Columbia, each with its own regulatory commission, political economy and consumer preferences. Some states may conclude that their customers are willing to accept a reliability premium – or a reliability discount – relative to a common PJM baseline. Under the current design, these preferences have no formal expression: Every state participates in a single uniform market that produces a single reliability outcome. A path toward geographic differentiation would require developing the technical and commercial framework for state-level or zone-level reliability procurement, shifting significant decision-making authority from the centralized PJM market to state regulators. That shift is architecturally complex and politically significant – the "leaning on neighbors" dynamic that characterizes current pool operations does not disappear cleanly, and the transition between a shared pool and a differentiated one requires careful boundary-setting.[23]

Customer-class differentiation. Within any geographic area, different types of load have fundamentally different values of reliability. A hospital's tolerance for interruption differs from that of an aluminum smelter or a cryptocurrency mining operation by orders of magnitude. The distribution utility system already manages this heterogeneity through tariff structures – interruptible service, curtailable contracts, demand response programs – and has done so for

---

[23] See also the discussion in closing section of Borenstein, James Bushnell, and Erin Mansur. "The economics of electricity reliability." Journal of Economic Perspectives 37.4 (2023), pp. 181–206:

[I]t may be possible to retreat from the axiomatic belief that reliability is a public good. Certainly within short operational time frames, shared responsibility for operating reserves will be necessary for the foreseeable future. However, over longer planning horizons it may be possible to identify control areas or individual load-serving entities that have failed to provide adequate resources and impose substantial penalties for their impact on the reliability of other customers. Ultimately it may become possible to interrupt only the customers of the inadequate service providers, although this would require being able to identify culpability for supply shortfalls in near real-time.

Thus, with emerging technologies and creative market design, it may be possible to allow individual load-serving entities to approach their resource acquisition according to their individual choices and beliefs about the market rather than through a standardized set of metrics and rules. Disagreements between local regulators and Independent System Operators about the likely effect of energy efficiency programs, intermittent supply, demand response, or even conventional generation can be put to the test by allowing local load-serving entities to make their choices, but also live with the consequences.

decades. The question for Path B is whether this logic can and should extend into the wholesale market structure: defining service quality tiers not just for the purposes of emergency curtailment, but as an organizing principle for how capacity is procured and allocated. Doing so would require moving beyond the current single-clearing-price auction to markets that explicitly price different reliability attributes separately – a significant market design undertaking.

Willingness-to-pay versus cost causation. These two principles for allocating reliability scarcity may lead to different outcomes and encode different values. A cost-causation framework prioritizes legacy load because those customers bore the historical cost of building the shared infrastructure; new loads that did not contribute to that investment should bear the residual curtailment risk. A willingness-to-pay framework would allow any customer – legacy or new – to purchase up to a higher reliability tier at a price reflecting the scarcity premium, regardless of historical contribution. Both are defensible; they reflect different judgments about whether grid reliability is primarily a shared public good (giving prior participants a claim on its proceeds) or primarily a private good (making current willingness to pay the relevant criterion). Different Path B implementations encode different answers to this question, often implicitly.

The Governance Challenge

None of these design questions can be resolved by PJM alone, and that is what makes Path B the most complex of the three paths, even if it is in some respects the most technically tractable. The decisions at stake – who bears the first-loss position during a grid emergency, how much reliability is owed to legacy ratepayers versus new large loads, whether states can opt into differentiated reliability standards – involve distributional consequences that require explicit governmental authorization and careful coordination across multiple institutions.

FERC has jurisdiction over the wholesale market rules that govern curtailment priority and capacity product design. State commissions have jurisdiction over the retail tariffs and service contracts through which customers would actually experience differentiated reliability. State legislatures and governors hold the authority to make the value judgments underlying any deliberate differentiation regime. Getting all of these institutions to act coherently on a question as novel as explicit reliability differentiation is a multiyear regulatory undertaking, not a near-term market design tweak.

This is not a reason to avoid Path B. It is a reason to begin framing the questions clearly now, so that when policymakers are forced to make choices a principled framework is available rather than improvised under emergency conditions. The Connect and Manage work underway in PJM's stakeholder processes is a meaningful start on one of the foundational questions. The deeper issues around geographic differentiation, customer-class tiers and the allocation principles governing scarcity rationing remain open, and this paper does not purport to resolve them. It raises them as the questions that Path B requires the region to eventually confront, deliberately and with appropriate governance.

Path C: Energy Market Transition (The "Shift the Revenue Recovery" Model)

Paths A and B both accept the capacity market as the primary vehicle for ensuring resource adequacy – they differ on whether that market should be made more financially durable (Path A) and/or restructured to apply differential reliability (Path B). Path C asks, and puts forward one potential answer to, a more fundamental question: *Should the capacity market remain the central organizing instrument at all?*

Under Path C, PJM would pursue a deliberate, phased shift of the revenue recovery function from the capacity market to the Energy and Ancillary Services (E&AS) markets. The ultimate objective of evolving the E&AS markets under this path is not merely to ensure they "work well with" the next iteration of the capacity market. The strategic goal would be to reduce the region's reliance on the capacity market over time.

The capacity market would progressively narrow – becoming a backstop for recovery of the remaining, shrinking "missing money" needed to support resource adequacy rather than a comprehensive annual revenue recovery mechanism – as E&AS markets are reformed to price scarcity more accurately and as forward energy contracts become the primary hedging instrument for Load Serving Entities. The thesis underpinning a potential decision to pursue Path C would be that a well-designed energy market, combined with enforced long-term forward contracting of that energy product, can achieve the investment coordination that capacity markets were designed to provide, without many of the political and administrative vulnerabilities that have plagued capacity markets in practice.

In this section, the rationale for and potential benefits of Path C are provided at a high level. Recognizing this path is a longer-term and more significant change from the current market structure in PJM, a more detailed overview of alternative approaches to implement this path is provided in Appendix C.

The Problem of Incomplete Scarcity Pricing

As established previously, the capacity market exists primarily to solve the "missing money" problem, that is, the revenue shortfall created when energy markets are unable to reflect the full value of reliability when such reliability is threatened. When operating reserves are depleted and the grid is under severe stress, economic principles dictate that prices should rise steeply to reflect the true marginal cost of maintaining system balance – sometimes, even including the cost of the load curtailment that was necessary to maintain energy balance. Instead, due to various factors, prices are frequently suppressed below efficient levels. Because the energy market is not allowed to fully compensate generators during times of extreme stress, the capacity market steps in to make them whole.

This incomplete scarcity pricing creates a cascade of negative effects:

- It results in the inefficient dispatch of resources during critical periods. The energy market provides tremendous value through the coordinated scheduling of resources in a least-cost manner, including storage and demand response. As we transition to a system with unprecedented growth in large, sophisticated data center loads and look to enable their demand-side flexibility, the efficiency value of the energy market fails to be fully realized when these resources are dispatched through administrative emergency actions rather than economically. This issue is further discussed in "The Hyperscale Paradigm Shift" in Section IV.
- It undermines the investment signal and financial viability for certain highly flexible resources. Resources critical for managing scarcity, such as fast-ramping units, multi-hour batteries or potentially price-sensitive demand that performs exceptionally well during real-time emergencies, may not receive the price signals necessary to justify their specific operational capabilities.
- It bloats the capacity market. The more we suppress prices in the energy market, the larger the "missing money" gap becomes, placing a larger burden on the capacity market to attract investment.

The profound opportunity to rethink this balance is explored in Path C.

The Theoretical Case for Returning Scarcity to the Energy Market

By allowing E&AS markets to accurately reflect market fundamentals during tight conditions, a greater share of total wholesale market revenues will shift from the forward capacity market to the real-time Energy and Ancillary Services markets. When energy prices are allowed to rise to efficient levels during scarcity, generators earn more of their required capital recovery precisely when they are providing the most value to the grid.

Consequently, as more of the “missing money” is recovered in the E&AS markets, the residual “missing money” gap that the capacity market must fill shrinks. This evolution represents a shift away from the administered capacity market and a return to energy market fundamentals. A well-functioning energy market can provide the revenue recovery that generators need to justify investment, making the capacity market largely redundant. While a capacity market will likely remain necessary during the necessarily gradual evolution under this path, its role can be appropriately right-sized.

William Hogan developed the theoretical case for high-price, energy-only markets in a series of papers from 2005 onward, arguing that the problem is not energy markets per se, but the combination of price suppression with inadequate scarcity pricing mechanisms.[24] Robert Wilson’s work on market design similarly concludes that energy markets can support adequate investment if scarcity rents are not systematically suppressed.

Peter Cramton and Steven Stoft extended this framework by proposing a specific instrument to make the transition tractable: the Reliability Option.[25] Rather than eliminating revenue certainty for generators, the Reliability Option converts the capacity market’s flat payment into a financial call option – generators receive a premium for committing to deliver energy at a specified strike price during scarcity events, but the spot market is free to clear at the true scarcity price, providing both the investment signal and consumer protection simultaneously.[26] It is worth noting that some ISOs with capacity markets, most notably ISO-NE, have deliberately moved to blur the line between energy and capacity, using designs that share many features with Reliability Options. Under the ISO-NE “Pay for Performance” capacity market design, a generator that sells capacity essentially sells their right to scarcity payments available to non-capacity resources during shortages. The capacity market designs of Ireland and Italy are examples of this where the obligation of committed capacity explicitly includes the Reliability Option for energy.

Frank Wolak has advanced a complementary argument focused on the demand side of the hedging problem. Wolak’s forward hedging framework proposes that Load Serving Entities would be required by regulation to maintain a minimum portfolio of long-term forward energy contracts – Power Purchase Agreements (PPAs), tolling agreements and financial energy hedges – covering a substantial share of their expected load, with the share declining as the delivery date approaches.[27] The logic is that the “credibility trap” described in this paper – where politically unacceptable scarcity prices invite intervention that undermines the investment signal – is primarily a problem of *unhedged load*, not a problem of the energy market itself. If load is hedged, high spot prices become an accounting transfer between the LSE’s energy position and its hedging portfolio rather than a rate shock to retail customers; governmental pressure to intervene evaporates.

Wolak’s framework has particular relevance to PJM’s current situation because it directly targets the mechanism of the credibility trap without requiring a redesign of the spot market. Mandatory hedging requirements could, in principle, be implemented in the near term as a structural reform to Path A, while the longer-run E&AS market reforms necessary for a full Path C transition are developed – making mandatory hedging a bridge strategy rather than a permanent destination.

---

[24] Hogan (2005, 2013), *op. cit.*; Robert B. Wilson, “Architecture of Power Markets,” *Econometrica* 70(4) (2002), pp. 1299–1340

[25] Here and elsewhere, the term Reliability Option is capitalized when used to refer to the concept of a “financial option,” and to clearly distinguish it from other “design options to achieve reliability.”

[26] Peter Cramton and Steven Stoft, “The Convergence of Market Designs for Adequate Generating Capacity,” *White Paper for the Electricity Oversight Board* (2006); Cramton and Stoft, “Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency,” *Utilities Policy* 16(3) (2008), pp. 194–201

[27] Frank A. Wolak, “Regulating Competition in Wholesale Electricity Supply,” in *Economic Regulation and Its Reform: What Have We Learned?*, ed. Nancy L. Rose (University of Chicago Press, 2014)

Recent academic work provides a rigorous framework for understanding why the *contractual form* of resource adequacy obligations matters – not just their existence. Shu and Mays (2022) demonstrate that mandatory capacity obligations crowd out the contracting arrangements that different technologies would naturally prefer: Baseload generators naturally sell long-term futures, peaking plants naturally sell call options and variable renewables naturally sell unit-contingent contracts tied to their actual output.[28] Forcing all resources into a common, constant-volume option instrument mismatches contractual form to physical risk profiles, generating risk premia that raise effective costs and distort investment incentives. Their preferred alternative – Standardized Fixed-Price Forward Contracts (SFPFCs) sold by a diversified portfolio of resource types – achieves near-optimal efficiency because the portfolio naturally hedges both price and load-shape risk, aligning with what balanced forward energy contracting already provides in bilateral markets. This analysis reinforces that the case for shifting revenue recovery toward long-term energy contracting is not merely that energy products are more legible to project finance (though they are), but that they better align contractual structure with the physical risk profiles of the diverse resource mix PJM will rely on over the coming decades.

Strengthening E&AS scarcity pricing also addresses, at its structural root, a fundamental limitation in the capacity market's ability to incentivize performance as the generation fleet diversifies. Zuo, Macey and Mays (2025) document a compounding challenge: As the resource mix shifts toward variable renewables and storage, the methodological demands of accurate accreditation increase precisely when the tools available to enforce it are most constrained.[29] Scarcity events are statistically rare by design, limiting sample sizes for accreditation studies; correlated supply failures – gas curtailments coinciding with suppressed solar and wind output during a winter polar event – confound average-based availability estimates; and non-performance penalties, the incentive complement to accreditation, are bounded by the creditworthiness ceiling that the same authors' earlier work identifies as a structural feature of any capacity obligation construct.

Full-strength E&AS scarcity pricing offers a structural resolution to this challenge: When generators earn revenue in proportion to their actual delivery during high-value scarcity hours, the energy market itself provides the performance incentive. A resource that fails to produce during a $10,000/MWh event bears the full revenue cost of that non-delivery; one that delivers earns the full scarcity premium. In the near term, this complements rather than replaces PJM's Capacity Performance (pay-for-performance) provisions: The administrative penalty framework continues to provide an explicit, contractual performance obligation, while E&AS scarcity pricing adds a direct market-based incentive layer alongside it. Over the longer run, however, as E&AS scarcity pricing matures and generators internalize the full revenue consequence of non-delivery, the marginal value of a separate administrative penalty structure diminishes – and the case for simplifying or eventually sunsetting redundant administrative overlays strengthens. The endpoint of this evolution, if pursued consistently, is a market where performance is enforced primarily by the energy market itself rather than by a parallel administrative construct.

By beginning to address incomplete energy market scarcity pricing, we align the financial incentives of the energy market with the physical needs of the grid – ultimately paving the way for a simpler, more resilient and more economically efficient wholesale market design.

---

[28] Han Shu and Jacob Mays, "Beyond Capacity: Contractual Form in Electricity Reliability Obligations," *Energy Economics* 126, 106943 (2022)

[29] Ke Xin Zuo, Joshua C. Macey, and Jacob Mays, "Revisiting Capacity Market Fundamentals," Working Paper, Cornell University / Yale Law School (January 2025)

High Scarcity Prices Need Not Mean Unaffordable Bills

A proposal to increase scarcity pricing in the energy market naturally invites immediate concerns about consumer affordability. If energy prices are allowed to rise to $10,000/MWh or higher during grid emergencies, policymakers will rightfully ask how ratepayers can survive such volatility.

The answer lies in understanding that the objective of complete (energy market) scarcity pricing is not to increase the total cost of the wholesale market, but to shift where the dollars necessary to build and operate generation are coming from. Furthermore, shifting revenue recovery to the energy market opens the door to superior hedging mechanisms that can shield consumers from volatility far better than the current construct.

The Affordability Equation: Shifting, Not Increasing, the Missing Money

The total cost to attract and retain generation in PJM is driven by the physical and financial fundamentals of building and running power plants (the Cost of New Entry). In our current design, because energy scarcity prices are suppressed, a significant portion of this required revenue must be collected through the capacity market.

If PJM and stakeholders reform the E&AS markets to allow higher, more accurate scarcity prices, generators will earn significantly more revenue during the hours when the grid is tightest. Mathematically, as a generator’s expected E&AS revenues increase, their “missing money” decreases. In a competitive market, this higher E&AS offset will drive down their capacity market offers, thereby reducing capacity clearing prices.

The total revenue the generator receives remains similar. It is simply rebalanced. Instead of consumers paying a high, flat “insurance premium” (capacity) every day of the year regardless of grid conditions, they pay a lower baseline premium and only pay high energy prices when the system is actually stressed.

Structural Advantages of Energy Market Hedging

Shifting value back to the energy market creates meaningful advantages for how Load Serving Entities manage their cost exposure – though those advantages are conditional and vary by contract tenor and other factors.

The capacity market is an administrative construct, highly vulnerable to “stroke-of-the-pen” regulatory risk. Rules regarding the VRR Curve shape and cap, Effective Load Carrying Capability (ELCC) accreditation and Market Seller Offer Caps (MSOC) change on a regulatory filing cycle; a long-term bilateral capacity contract is priced against administrative parameters that may look entirely different by the time the contract matures.

A megawatt-hour of energy, by contrast, is a physical, standardized commodity. Forward energy contracts – power purchase agreements, tolling agreements and financial energy swaps – are priced against observable supply and demand fundamentals and trade in markets with established commodity price forecasting methodologies. At medium tenors (roughly 3–7 years), bilateral forward energy markets are meaningfully more liquid and more legible to project finance than their capacity market counterparts. For this tenor range, energy contracts are a structurally more durable hedging instrument than bilateral capacity contracts.

That said, energy forward markets are not uniformly deep across all tenors. Open interest in energy financial contracts thins materially beyond 5–7 years; long-dated bilateral physical contracts are typically indexed to floating prices unless fuel revenue streams can be simultaneously fixed. This liquidity gradient is precisely why the spectrum of centralized instruments – Reliability Options, CIS revenue floor/ceiling structures and the Australian ESEM – exist: They address the longer-tenor segment where bilateral market depth cannot do so on its own. Forward energy hedging is not a complete substitute for market design; it is a more suitable instrument than the current administrative capacity market for the tenors where bilateral markets function, and centralized structures are needed for the tenors where they do not. Various approaches to implement forward hedging are explored in more detail in Appendix C.

Trade-Offs To Consider

Path C is not presented as the near-term solution to PJM's resource adequacy challenge. It is presented as a coherent long-run direction that the region might commit to – one with genuine advantages over the long run but with significant prerequisites and transition risks. These include:

- Path C requires substantial E&AS market reforms as a necessary precondition. Energy price caps are raised over time, and the market's scarcity pricing mechanism must be robust enough to survive the public pressures that have historically driven intervention by government. These reforms are non-trivial and contested.
- A transition period in which capacity market signals have been weakened but E&AS market reforms are not yet complete could open a window of elevated reliability risk. Conversely, a transition period in which scarcity pricing has been strengthened but the capacity market has not yet receded raises affordability concerns. Managing a transition under this path requires careful sequencing.
- Mandatory LSE hedging requirements require regulatory authority that may need FERC action and state cooperation, particularly for default service providers regulated at the state level. This creates jurisdictional complexity.
- The Path C endpoint – a fully energy-market-centric system with mandatory forward energy hedging – is a longer-term-horizon commitment based on a conscious fundamental shift in philosophy, not a 2027 deliverable. In the near term, the most immediate contribution of Path C thinking is to validate and expand the mandatory hedging elements of Path A, while building the E&AS market reforms that make a longer-run transition feasible.

Relationship to Paths A and B

The three paths are not fully independent design silos. The mandatory hedging elements of Path A are structurally consistent with Path C: requiring LSEs to hold long-term forward energy contracts is exactly the mechanism that the Path C design literature (Wolak, Australia CIS) proposes to protect consumers as the capacity market's role is progressively reduced. Implementing meaningful hedging requirements under Path A could therefore function simultaneously as a near-term stabilization measure and a first step in a longer-run Path C transition.

Path B, by contrast, involves a different organizational principle – accepting reliability differentiation rather than working to universalize reliable investment signals – and is less naturally complementary to Path C. However, even under Path B, all but the most price-sensitive loads would benefit from energy market hedging instruments, and E&AS market reform remains a necessary complement.

The practical implication is that PJM may be operating on multiple tracks simultaneously: implementing Path A hedging reforms in the near term (2026–2029), piloting Path B differentiation frameworks for large new load (2027–2030), and developing the E&AS market reforms that would make a longer-run Path C transition feasible (2028 and beyond). These are not mutually exclusive commitments – but clarity about the long-run directional goal matters for market design choices made today.

Other Supporting Capacity Market Design Reforms

While not directly responsive to the primary strains on the region's resource adequacy construct today, there are additional reforms to the capacity market design that are significant and deserve further discussion and consideration under multiple paths forward and/or as part of a transition path. This section discusses the following three potential areas of reform that fall into this category:

- A sub-annual or seasonal market design where capacity is separately procured and priced for different periods of the year
- Reevaluation of the auction forward period to either align with current-day construction timelines, or a consideration of a "prompt" auction design where the BRA (or similar residual procurement) would be held closer to the start of the delivery year rather than three years forward
- Continued enhancements to risk modeling and ELCC accreditation of capacity

In addition to these capacity market reforms, the need for significant reforms in the Energy and Ancillary Services markets is discussed in the section that follows.

Sub-Annual Capacity Market Design

PJM and its stakeholders have recognized the potential reliability and efficiency benefits of a sub-annual capacity market construct for some years now. In late 2025, an independent consultant (Analysis Group) was hired to conduct a comprehensive review of those potential benefits and to evaluate the trade-offs of different design considerations under a sub-annual framework.

The consultant provided their final report in January 2026 with the recommendation that PJM and its stakeholders pursue the development of a sub-annual (two season) capacity market design and highlighted the following potential benefits in moving from the existing annual construct to a sub-annual market:

- Improved capacity pricing to achieve better short-term allocation of resources and long-term investment in capacity resources
- Improved accounting of resource and system features that vary across sub-annual periods
- Improved alignment of resource costs and obligations and resource compensation with the services provided
- Reduced year-to-year variability in resource accreditation and greater flexibility to adapt the RPM to changing market circumstances

The consultant also discussed the costs and trade-offs associated with various sub-annual market structures that would need to be more fully vetted and considered if pursuing the sub-annual design.

Reevaluation of Forward Capacity Auction Timeline

One of the key design elements at the inception of RPM was the three-year forward procurement horizon, which was designed to enable more direct competition between new entry and existing assets as well as provide a forward price signal of capacity needs.

However, those benefits and the associated trade-offs that come with a three-year forward procurement have come into question more recently as system conditions have evolved.

First, the project timeline associated with developing a new combustion turbine (the reference technology in the capacity market) has increased significantly under supply chain constraints and other factors, greatly extending the time needed to bring a new unit online from what's been historically required. As such, the ability of the three-year forward auction design to provide its intended benefit of direct new entry competition is severely impaired at this time, particularly for the reference technology, and it is unclear if or when it may be regained.

Second, load forecast uncertainty has increased significantly in recent years. Historically, forecast uncertainty was primarily driven by weather volatility and macroeconomic trends, but the rapid expansion of data centers has introduced a new, structural form of uncertainty. This increases the uncertainty of capacity requirements for the system, particularly further out into the future, impacting the balance of trade-offs to consider when selecting a forward period. All else equal, the prompt auction design helps address that uncertainty both from the load forecast perspective in having better line of sight into the level of data centers that have materialized, and a better projection of the resource mix that will exist on the system to use in the resource adequacy studies. This, in turn, may allow for more robust representations of supply, demand and clearing outcomes in the auction.

The details of moving to a prompt auction would need to be developed in conjunction with any solutions under the Path A framework that seek longer-term commitments of capacity. Additionally, the benefits noted above would need to be carefully weighed against potential drawbacks in moving to the prompt auction. This would include reevaluating the shortcomings observed under the short-term capacity credit market prior to the RPM that had led to a forward market design, and if or how those may be addressed through other structural reforms.

ELCC Enhancements

Accurate resource adequacy analysis and capacity accreditation is essential for determining the total capacity requirements needed to satisfy PJM's 1-in-10 LOLE reliability standard, and to ensure that resources are fairly compensated for the resource adequacy value they provide to the grid.

In 2024, the Commission accepted PJM's proposal to enhance its resource adequacy analysis, shifting to an hourly framework that more comprehensively assessed loss-of-load risk on the system across the year and better captured the relationship between extreme weather conditions, load profiles and correlated outage risks of generation. The Commission also approved the adoption of a marginal ELCC approach for capacity accreditation, providing a more accurate representation of the resource adequacy contribution of different generation classes and individual units.

These changes correspond to a growing trend in the industry where more sophisticated and complex modeling of resource adequacy is required, with many ISOs shifting to more granular resource adequacy studies and capacity accreditation approaches like marginal ELCC.[30] As the level of complexity has increased, PJM has worked with members on continued improvements to model transparency and further understanding of the ELCC analysis and accreditation,[31] as well as worked with stakeholders on continued enhancements to the analysis.

---

[30] ESIG Report on Capacity Accreditation (PDF)

[31] Extensive documentation describing the resource adequacy analysis and ELCC accreditation, as well as the underlying data used within the analysis, are provided on PJM's Effective Load Carrying Capability web page.

At the direction of the PJM Board, PJM retained an independent consultant in late 2025 with expertise in this area (E3 – Energy and Environmental Economics, Inc.) to perform a comprehensive review of the PJM resource adequacy and capacity accreditation analyses for alignment with industry best practices and to identify potential areas for improvement. In totality, the review found that PJM's model was fit for purpose, sophisticated and aligned with industry best practices. They also identified several opportunities for PJM to consider making improvements. Their review and opportunities for enhancements were detailed in a published report[32] and warrant further evaluation with PJM stakeholders as we explore paths forward for the capacity market design.

E&AS Market Reforms Necessary Under Any Path

Activating data center flexibility – and more broadly, ensuring that the E&AS markets can efficiently price reliability – requires foundational reforms to PJM's reserve markets that are overdue regardless of which capacity market path the region ultimately chooses. Paths A, B and C all require a reserve market infrastructure that accurately values operational flexibility and sends price signals sufficient to attract and retain the resources the system needs.

PJM has been working with stakeholders in the Reserve Certainty Senior Task Force (RCSTF) to explore exactly these reforms. PJM considers this package necessary to value the essential reliability services the system needs to continue operating reliably – and critically, this work should proceed in parallel with, not dependent upon, the resolution of the longer-horizon capacity market path question.[33]

The Operational Problem

The urgency for these reforms is demonstrated by current operating conditions. In 2023, PJM's hour-ahead net-load forecast error exceeded the largest contingency in more than 130 hours – meaning that the primary tool the market uses to maintain contingency reserves was already insufficient to address the variability the system was experiencing. As the resource mix continues to evolve, this challenge will compound. Net-load forecast uncertainty will increasingly dwarf the contingency risk that current reserve products were designed to manage.

The existing reserve market structure was designed for a fleet dominated by dispatchable thermal units with predictable availability patterns. It relies on a single dominant concept – contingency reserves for the largest unit loss – and calibrates penalty prices against 2007 data. The current Synchronized Reserve ORDC penalty of $850/MWh has not been updated to reflect either current cost conditions or the much higher cost of load shed events (the Value of Lost Load).

Additionally, PJM routinely commits resources out of market and day ahead to meet reliability needs that the current market cannot procure transparently. Consistent, predictable, out-of-market actions are a signal that the market design does not adequately reflect operational requirements – and they create cost-recovery ambiguity for the resources providing those services.

The RCSTF Reform Package

PJM's RCSTF package proposes six categories of reform to address these gaps:

1. **Day-Ahead Scheduling Reserves (DASR).** A new product to procure reserves day ahead to address the uncertainty in load, solar and wind forecasts and generator performance risk. Unlike current products that focus on contingency events, DASR is designed to manage the probabilistic

---

[32] E3's Report on PJM's RRS/ELCC Model Evaluation (PDF)

[33] PJM Interconnection, PJM's Position on Challenges and Solutions for Long-Term Reserve Certainty Reforms, Reserve Certainty Senior Task Force, April 6, 2026, draft, pp. 4

forecast errors that drive the bulk of near-term operational risk. The ORDC for DASR is calibrated at $50/MWh – a modest signal for day-ahead commitment decisions – with quantities calculated at different percentile risk levels based on historical data.

2. **Updated Synchronized Reserves with recalibrated ORDC.** The current $850/MWh Synchronized Reserve ORDC penalty would be raised to $2,100/MWh – a level designed to ensure that all more cost-effective measures are exhausted before resorting to emergency load shed. A locational procurement approach using actual distribution factors would replace the current RTO-wide structure, improving the market's ability to manage constrained areas.
3. **New Ramp/Uncertainty Reserve (RUR) products.** Two new products – a 10-minute RUR and a 30-minute RUR – would explicitly manage net-load forecast uncertainty and expected ramping needs. These are the products most directly relevant to activating data center and other large-load flexibility: the 10-minute RUR's ORDC is anchored at $1,000/MWh at the expected ramp point, rising to approximately $1,900/MWh at tighter reserve levels. At these price levels, large flexible loads face a genuine economic incentive to provide services. Other ISOs, including CAISO, MISO and NYISO, have already implemented analogous products.

4. **Redefined 30-Minute Reserves.** The existing secondary reserve product would be explicitly redefined to address the need to backfill 10-minute contingency reserves following a contingency event – clarifying its purpose and nesting its procurement requirement within the RUR framework.

5. **Energy Gap Reserves.** A new day-ahead-only product for elevated-risk winter days (November through March) to address the gap between the day-ahead load forecast and cleared physical supply. Historical energy gap distributions suggest a 90th percentile gap of approximately 4,500 MW; the product would procure reserves on a cost-benefit basis along a demand curve calibrated to historical gap data.

6. **Reserve Offer Reforms.** Allow resources to recover legitimate availability costs through reserve offer structures rather than requiring near-zero bids. The current requirement to offer reserves at or near zero misaligns market incentives with operational reality and discourages resources from maintaining the operational readiness the system needs.

Why This Cannot Wait

PJM's RCSTF package explicitly identifies urgency: These reforms will take time to design and implement, and their absence before operational issues materialize creates both reliability risk and cost risk. The market reforms needed to properly price and procure ramp capability, uncertainty management and operational flexibility are not a luxury – they are the foundational infrastructure on which data center flexibility, demand response and Path C-enabling E&AS reforms all depend.

Importantly, the RCSTF reforms are not contingent on any particular capacity market design outcome. They improve the efficiency and reliability of the real-time market regardless of whether PJM eventually moves toward Path A (mandatory hedging with refined RPM), Path B (differential reliability) or Path C (progressive energy market transition). Getting these reforms right early reduces the cost of transition under any path and expands the set of tools available when those path-specific choices are made.

The Hyperscale Paradigm Shift

The coming wave of macro-load is fundamentally different from the residential and commercial load that defined the grid of the 20th century. Hyperscale data centers are unprecedented in their scale, with individual facilities approaching the size of traditional power plants. And they are managed by highly sophisticated entities that possess a deep understanding of energy markets, advanced operational capabilities and multiple levers of inherent flexibility:

Workload flexibility. AI compute workloads divide into two categories with fundamentally different flexibility profiles. AI *inference* – the real-time use of trained models to answer queries – is latency-sensitive and effectively firm load. AI *training* – the process of building models from large datasets – is throughput-oriented and has natural flexibility. Google estimates that 40% of its AI energy use comes from the training phase.[34] Training workloads have built-in “flex points” at checkpoint intervals, gradient accumulation windows and parallelism strategy boundaries that can tolerate short slowdowns or pauses without compromising the training run’s ultimate output. These are not theoretical flexibilities; they represent real operating practices that sophisticated operators already manage to allow them to reschedule lower urgency computational needs while protecting the highest priority workloads.

On-site generation and storage. Data centers have historically been built with diesel backup generators capable of running the full facility load during grid outages. As battery energy storage becomes cost-competitive, replacing those diesel generators with lithium-ion batteries creates a resource that is not merely a passive backup – it may be configured to be able to actively discharge to the grid during peak demand hours, providing both the data center’s IT reliability and a grid service simultaneously.

Advanced load control architectures. Next-generation data center facilities are being designed with granular, software-controlled load management that can reduce facility demand on a graduated rather than binary basis. This is qualitatively different from traditional industrial curtailment (which typically means stopping a process entirely): These systems can reduce power by 10%, 20% or 30% in seconds through integrated, coordinated control of workload scheduling flexibility and onsite storage discharge.[35]

But accessing even a portion of that demand flexibility economically through the energy or reserve markets will require evolving those markets. Current market caps and rules, designed to protect passive consumers, inadvertently suppress the signals needed to activate active consumers. Analysis of the cost of this flexibility reveals that while cryptocurrency miners already regularly suspend operations when electricity prices exceed operational profitability, data centers engaged in AI computations may see little economic incentive to reduce demand based solely on modest price signals.[36] At current PJM energy price cap levels (roughly $3,700/MWh), the economics of curtailing compute operations are generally unattractive for most AI workloads.

At significantly higher price levels – in the $10,000/MWh range, closer to the true value of reliability – the calculation changes materially. Recent analytical work by Hans Royal introduces the concept of the *Compute Heat Rate* (CHR) – the maximum electricity price at which a given AI workload remains profitable, a demand-side analog to the gas heat

---

[34] Sidewalk Infrastructure Partners (SIP), *Data Center Flexibility: A Call to Action* (March 2024)

[35] For further discussion of data center flexibility capabilities and demonstrations, see:

NVIDIA and Emerald AI, Press Release, CERAWeek 2026 (March 2026),

Sidewalk Infrastructure Partners (SIP), *Data Center Flexibility: A Call to Action* (March 2024),

Emerald AI, EPRI, National Grid, and Nebius, *Power-Flexible AI Factories: A UK-First Demonstration of Grid-Responsive AI Infrastructure* (March 2026), pp. 2, 5–6. The demonstration was conducted as part of EPRI’s DCFlex initiative.

[36] RMI, “How Data Centers Can Set the Stage for Larger Loads to Come” (May 2024)

rate familiar from resource dispatch modeling.[37] Royal's analysis estimates a blended average CHR across workload types of approximately \$6,300/MWh – roughly 127 times the current wholesale average but squarely within the range that could be accessed under achievable scarcity pricing reforms. At a \$10,000/MWh scarcity price, workloads with the lowest CHRs – commodity AI tasks and enterprise contracted jobs with CHRs in the \$800–\$1,270/MWh range – become economic to curtail. Mid-tier inference workloads (CHR approximately \$8,000/MWh) approach their threshold. Frontier AI operations (CHR exceeding \$50,000/MWh) continue uninterrupted. This is precisely the graduated, price-elastic demand response the grid needs: not a binary shutdown, but a tiered response where the most flexible compute responds first, in proportion to price. The RCSTF's proposed Ramp/Uncertainty Reserve (RUR) ORDC, anchored at \$1,000/MWh rising to ~\$1,900/MWh, begins to create the signal necessary to activate this flexibility; a higher cap ceiling would reach further into the CHR distribution.

If we can activate the demand-side flexibility of large loads economically, we can transition from relying on administrative emergency actions to using economic rationing, fundamentally transforming how the grid manages scarcity. In a modernized E&AS framework, when the system becomes constrained, the market does not "break." Instead, the price rises to a level that reflects the true value of the scarce energy. At a certain price point, it becomes economically rational for a sophisticated data center to seamlessly transition to its on-site backup generation, discharge its batteries or throttle its compute load. Just as importantly, it becomes economically rational for these new loads to invest in more sophisticated demand management approaches and more efficient on-site backup generation that could operate more frequently, in periods beyond the most severe grid emergencies. As they do so, the vertical demand curve bends – demand voluntarily decreases, supply equals demand and the market clears efficiently.

The uniqueness of this type of load, combined with the scale of load growth, present a generational opportunity to rethink not only the solutions to the resource adequacy problem (as presented in Paths A, B and C) but also the problem itself. As discussed in Section II, the resource adequacy problem stems from a combination of insufficient demand elasticity and the inability of the system operator to curtail specific loads. Both of these structural characteristics, taken as bedrock assumptions when the current resource adequacy framework was established, may be fundamentally altered by the arrival of hyperscale data center loads. Paradoxically, this could mean that the influx of the very loads contributing to the resource adequacy problem today will help to shrink the relative scale of the missing money problem over time. It is incumbent on PJM and its stakeholders to revisit the foundational assumptions that led to the current problem definition as we explore all options for its solution.

---

[37] Hans Royal, "The Compute Heat Rate (CHR): Framework & Methodology" (February/March 2026)

V. Next Steps: Toward a Regional Deliberation

This paper does not recommend a path. That is a deliberate choice, not a failure of resolve.

The three paths described above are not market design proposals ready for implementation. They are a structured framing of a set of choices – about investment risk, consumer protection, reliability standards and market architecture – that the PJM region must eventually make explicitly or will make implicitly by doing nothing. PJM's role is to provide education, analysis and its expert recommendation to the discussion so that all parties involved in the discussion move forward with a clear understanding of what is at stake. As described herein, not all components of the solution are under PJM's control, and therefore collaboration across the entire set of stakeholders is necessary to get to a durable solution.

What PJM can commit to is a clear sequence of near-term actions that advance the region's options regardless of which path policymakers ultimately choose.

1 Complete the RCSTF package.

Move reforms through the regulatory process.

2 Develop the Connect and Manage framework.

Operationalize resource adequacy principles.

3 Begin the forward contracting policy work.

Explore the legal framework.

4 Engage with stakeholders.

Facilitate a structured stakeholder dialogue.

First, complete the RCSTF package. The E&AS market reforms currently under deliberation in the PJM stakeholder process – recalibrated reserve products, updated reserve demand curves and the Ramp/Uncertainty Reserve – are necessary regardless of which capacity market path the region follows. They address operational gaps that exist today, independent of the long-run architecture question. PJM should move these reforms through the regulatory process without waiting for the broader path decisions to be resolved.

Second, develop the Connect and Manage framework in the stakeholder process. The curtailment priority work underway in PJM's stakeholder processes is the nearest-term operationalization of the resource adequacy principles at stake. Getting this framework right – technically sound, legally defensible and fair to legacy customers – is a prerequisite for both Path B and other E&AS reforms to better integrate large load flexibility.

Third, begin the forward contracting policy work. The hedging instruments described in Paths A and C – from voluntary long-term bilateral contracts through mandatory LSE obligations to ISO-operated forward energy markets – require regulatory and, in some cases, legislative action at the state level to be viable at scale. That process takes years. PJM should work with state commissions and legislators now to explore the potential legal framework, even as the underlying path decisions remain open.

Fourth, engage systematically on the foundational questions both internally and with stakeholders. The questions that differentiate Paths A, B and C – how much reliability risk should consumers bear versus investors; whether the "Shared Reliability" compact can be maintained; whether forward energy contracting can replace administrative capacity procurement – are ultimately policy and political questions, not engineering questions. PJM will facilitate a structured stakeholder dialogue on these questions through 2026, with the goal of developing a sufficient regional consensus to move toward concrete market design changes.

The urgency is real. The supply-demand gap the region faces is substantial – it is visible in current interconnection queues, retirement notices and reserve margin forecasts. The time available to make these decisions deliberately, before operational conditions force them by default, is measured in years, not decades. PJM's commitment is to ensure that the region uses that time well.

Appendix A. History of PJM's Capacity Construct

This appendix provides a historical overview of how PJM's capacity construct has evolved over time into the market design in place today. It reviews the predecessor capacity constructs and drivers that led to the current market framework (the Reliability Pricing Model or "RPM"), the key design elements when RPM was implemented in 2007, and some of the notable modifications to the market design over the past two decades.

A foundational element that has remained throughout this evolution is the inherent value of the "PJM pool." PJM members, and ultimately the consumers across the footprint, have long benefited through the sharing of available resources to reliably serve load. The aggregation of resources and the diversity in regional load profiles across PJM's large geographic footprint helps mitigate risk, and it allows the system to reliably serve loads with a much lower reserve margin than what any single service territory would require in isolation. This dynamic has provided significant reliability value and savings for customers in the PJM footprint over the years.

Furthermore, PJM has long relied on resource adequacy standards and a capacity construct, or mechanism that sets capacity requirements on LSEs, to support reliable service to loads. Since the 1960s, PJM has used probabilistic methods to determine the pool-wide reserve margin and capacity needed to meet the one day in 10 years LOLE standard of the system. The analysis accounts for uncertainty in factors affecting reliability, such as generator performance and load characteristics, to determine the total installed capacity needed to satisfy the region's LOLE criterion. This targeted capacity level is expressed in terms of a percent above peak load (Installed Reserve Margin or "IRM") and provides the baseline for setting the capacity requirements of LSEs in the pool. In addition to the pool-wide resource adequacy standard, PJM has long used deliverability standards to test the transmission system's ability to deliver energy from, and to, various parts of the footprint to help ensure locational reliability.

Administrative Reserve Sharing (1974–1999)

Prior to the shift to competitive capacity markets, resource adequacy was managed through the PJM Interconnection Agreement where PJM members (utilities) agreed to share reserves to meet the pool-wide requirement.

From 1974 to 1999, the capacity construct imposed a two-year forward annual capacity obligation. The total capacity requirement for the pool was allocated among all member utilities, and each utility was required to demonstrate that it had, or would have, sufficient installed capacity to meet its load and reserve margin obligations two years ahead of the delivery year. Any utility that failed to meet its capacity obligation was assessed a capacity deficiency rate based on the estimated Cost of New Entry (CONE) of a combustion turbine.

While not a centralized capacity market, the deficiency rate was akin to a price signal in providing a financial incentive for each utility to meet its "equitable" share of the pool requirement and avoid paying the penalty, helping to deter any single entity from excessively leaning on the rest of the system for reliability.

Deregulation and the Capacity Credit Market (1999–2006)

The late 1990s ushered in the advent of deregulation and retail choice across several states in the PJM footprint. Deregulation involved the "unbundling" of the traditional electric utility, a process that separated the ownership of generation assets from the delivery of power through transmission and distribution lines. This shift enabled retail choice where individual customers can choose their own electric supplier rather than being tethered to the local utility's generation. Consequently, the responsibility for securing sufficient capacity to meet reliability standards shifted from the traditional utility members to various competitive LSEs in these regions.

To accommodate the introduction of retail choice, the two-year forward annual capacity obligation was replaced in 1999 with daily capacity obligations and the implementation of the daily and monthly Capacity Credit Market. The shift to short-term obligations facilitated retail competition by ensuring that the capacity obligation associated with a particular load (e.g., a retail customer) would promptly shift from one LSE to another when the customer changed suppliers. The Capacity Credit Market was designed to provide a competitive market where LSEs could trade capacity credits daily to meet the capacity obligations of the load they served and avoid a deficiency penalty.

While the Capacity Credit Market was well suited to facilitate retail choice, it ultimately proved inadequate for ensuring long-term resource adequacy in PJM. This dysfunction was driven by three primary issues:

1. No recognition of locational value. The price did not vary by location or reflect that capacity is not always universally deliverable throughout PJM. Thus, the price failed to signal the value of capacity or incent new generation in the locations that needed it most, threatening reliability. This issue was observed at the time in the eastern portion of the PJM region, where high load growth and increasing rates of generation retirement were driving reliability violations, and the need to retain existing units through out-of-market reliability-must-run (RMR) contracts.

2. Price volatility that failed to provide sufficient revenues to cover the cost of new construction or retain existing assets. The daily and monthly Capacity Credit Market prices were very volatile, with daily prices at or near zero for most of the five years between 2000 through 2004 with occasional spikes. The short-term nature of the market and single-value deficiency rate contributed to the boom-bust cycles and volatility in prices that failed to spur the needed investment.

3. No long-term forward commitment or forward price signals. The market rules only required that capacity resources be committed for as short as one day, with limited incentive for longer-term commitments. While this was designed to accommodate short-term competitive load switching under retail choice, the short-term market ultimately failed to demonstrate the capability to sustain long-term generation investment.

Reliability Pricing Model (2007 to Present)

In 2007, PJM transitioned to the current capacity market framework (RPM) with the overarching goal of aligning capacity pricing with system reliability requirements and providing transparent information to all market participants far enough in advance for actionable response to the information. The RPM construct was designed to address the shortcomings of its predecessor, shifting from a short-term procurement to a longer-term, forward-looking market design that signaled when and where investment was needed on the system and helped preserve reliability over the long term.

Key Design Elements of RPM

Three-Year Forward Procurement and Price Signal

The timeline of a three-year forward procurement of capacity resources and price signal was designed to provide a meaningful opportunity for participation of new entrants in the auction process, and to set sufficiently forward commitments of resources to help inform and coordinate the build-out of transmission on the system.

Locational Capacity Pricing

Locational price signals for capacity were designed to account for the physical transmission limitations of the system and to recognize the differentiated value of capacity across the footprint. This allowed import constrained areas to reflect the higher value of capacity and incent retention of existing resources and investment in the area.

Sloped Demand Curve

A downward-sloping demand curve (Variable Resource Requirement or VRR Curve) in the auction was designed to capture the incremental value of capacity at varying reserve margin levels, including those beyond the target reliability requirement, and provide sufficient revenues to support investment on the system over time. Additionally, the sloped curve provided a more stable price signal relative to the boom-bust cycle of the prior construct that relied on a single-value or vertical demand curve, helping to reduce investor risk and promote investment.

Facilitation of Retail Choice

The market design continued to facilitate retail choice in restructured states by enabling the allocation of capacity procured in the auctions and associated reliability charges to LSEs within zones be updated on a daily basis during the delivery year to promptly track customers and associated loads switching to an alternative supplier.

Market Structure Overview

Under the RPM, a series of auctions are conducted by PJM to secure capacity commitments on behalf of LSEs and to establish corresponding reliability charges for each delivery year.

RPM Auctions

The first and primary auction, the Base Residual Auction (BRA), is conducted three years in advance of the delivery year and is used to commit resources needed by LSEs after taking account of all self-supplied and bilaterally contracted resources. Market sellers provide competitive offers for new or existing capacity resources in the auction, which are cleared against the sloped demand curves of the RTO and LDAs on behalf of all LSEs. Subsequent incremental auctions are then run to provide a mechanism for market sellers to adjust their capacity positions, and for PJM to purchase additional capacity or release previously procured capacity, as the load forecast and reliability requirements for the delivery year get updated. This structure was designed to support reliability and new entrant competition by holding the BRA well in advance of the delivery year, while also allowing for flexibility with incremental auctions accounting for updated load forecasts and resource information leading up to the delivery year.

The results of RPM auctions set the clearing prices paid during the delivery year to capacity resources that clear the market and the reliability charges that will be borne by LSEs during the delivery year on behalf of their loads.

Support for Bilateral Contracting: *At its inception, the BRA was designed to be a “residual” procurement of capacity, and through providing transparent reference prices, the market would facilitate and promote self-supply and bilateral arrangements for LSEs to hedge against the RPM auction prices. Additionally, the design provided a set of standard bilateral transactions for market participants to use and required PJM to support electronic tools for reporting and tracking of those transactions, as well as an electronic bulletin board for prospective buyers and sellers to transact outside the RPM auctions.*

LSE Capacity Obligations and Reliability Charges

The LSEs serving RPM load in PJM are allocated a share of the total capacity procured in the RPM auctions for the delivery year and are responsible for paying the corresponding allocated cost of procuring it. The allocation of procured capacity is a two-step process. First, the total cleared capacity is allocated to individual zones based on each zone's forecasted peak load relative to the total RPM forecasted peak load. It is then the responsibility of EDCs within the zones (subject to regulatory oversight) to determine how the zone's allocated share of capacity is further allocated to individual customers and LSEs within their service territories.

The LSEs' reliability charges within a zone are then based on their allocated share of the capacity requirement and the zonal capacity price (based on the locational clearing prices of the RPM auctions) for such delivery year.

Fixed Resource Requirement (FRR) Alternative

The capacity construct also provided an opportunity for certain LSEs (investor-owned utilities, electric cooperatives and public power entities) to voluntarily opt out of the RPM auctions and instead meet the capacity requirements of their load through a long-term resource plan consisting of self-supply and bilaterally contracted resources. An FRR entity is required to meet the entire capacity obligation of their load, including load growth, for the applicable FRR service area. The capacity obligation of FRR load would be fixed at a quantity based on the forecasted peak load and IRM for the delivery year rather than a variable requirement as it is in RPM.

Performance Assessments

To help ensure committed Capacity Resources delivered on their forward obligations and reliability value during the delivery year, a set of performance assessments and testing requirements were implemented with the market design. These included deficiency assessments for resources that failed to be operational or provide sufficient capacity value to meet their forward commitment, testing requirements to confirm the rated capability of plants (and curtailment capability of demand response), as well as performance assessments to help ensure resources were available or capable of performing during peak periods of the delivery year. Any shortfalls assessed would then be subject to financial penalties.

Notable Modifications Over Time

While the high-level design objectives and elements of RPM have largely remained the same over the past two decades, there have been several significant modifications to the construct since its inception. The following provides a short history of certain notable modifications.

Capacity Performance

Following the reliability challenges experienced during the 2014 Polar Vortex, PJM transitioned to the Capacity Performance (pay-for-performance) market design to help incentivize resource performance during times of system emergency. This design set stricter performance obligations on committed capacity resources with steep financial penalties assessed on resources that failed to perform during the defined emergencies (i.e., Performance Assessment Intervals, or PAIs), and bonus payments paid to resources that performance beyond their committed obligation share.

The design was a shift from an assessment of broad availability over many "peak hours" for generation in the delivery year to one focused on delivered performance during the periods of greatest reliability need, thereby encouraging investment to improve resource performance during those times. Following the Winter Storm Elliott event in

December 2022 that saw $1.8 billion in assessed non-performance penalties, the Capacity Performance rules were modified to both limit the triggering conditions of PAIs, as well as adjust the maximum annual exposure of non-performance penalties for generation owners.

Minimum Offer Price Rule

The Minimum Offer Price Rule (MOPR) was included in the RPM market design to help address buyer-side market power concerns and ensure competitive clearing outcomes and prices. However, the design of MOPR has been a long-standing area of debate, often driven by the question of if, and how, state subsidies should be accounted for when clearing the wholesale market and setting competitive price levels.

The design of MOPR has been modified a number of times since RPM's inception, and in just the past decade, FERC's position on MOPR has shifted quite significantly. In 2019, following a delay of the RPM auctions as the MOPR issue was disputed at FERC, FERC issued an order to broadly expand the application of MOPR ("Expanded MOPR") to almost all new and existing resources receiving some form of state subsidy, effectively requiring that those resources offer at an administratively determined "floor" price representative of their net avoidable costs absent the subsidy.

Then, in 2021, these rules were largely reversed as PJM filed and had accepted by FERC a "Focused MOPR" that significantly limited the application of MOPR on resources under state subsidies and instead focused on concerns around buyer-side market power.

Marginal Effective Load Carrying Capability ELCC Accreditation

In early 2024, FERC accepted PJM's proposed reforms to use marginal Effective Load Carrying Capability (ELCC) accreditation for capacity and an enhanced risk model that better captured the relationship between extreme weather patterns, load profiles and resource performance on the system. Marginal ELCC replaced the prior Equivalent Force Outage Rates (EFORd) approach for assessing the capacity value of thermal generation, shifting to a method that accounted for correlated outage risks and reflected the incremental resource adequacy value a resource provided to the system. Additionally, accreditation for intermittent resources was changed from Average ELCC to Marginal ELCC at this time.

Market Successes

While the RPM market design is straining under the current environment of unprecedented demand growth outpacing supply, the past two decades have largely seen it accomplish what it was designed to do: provide a competitive procurement of capacity and forward price signals that have helped coordinate the efficient entry and exit of resources on the system to meet PJM's resource adequacy targets. These successes include:

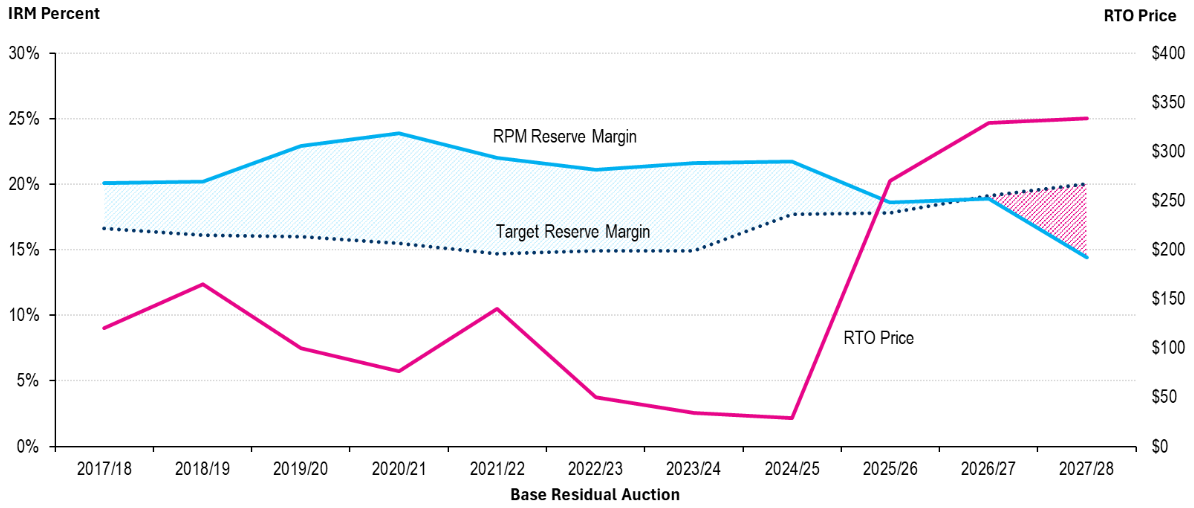
1. **Supported PJM's Ability To Maintain Resource Adequacy.** The capacity market design alone cannot guarantee resource adequacy, but since its inception in 2007 and up through the current delivery year, the RPM market has supported the efficient entry and exit of resources on the system while maintaining sufficient reserves to meet system and locational reliability requirements. In its early years, the RPM market helped address the reliability violations facing the region at its start, credited with attracting and retaining a significant amount of resources across the first five BRAs.[38] For most of the 2010s and early 2020s, the RPM market has secured reserve margins that have often exceeded the targeted level and done so at a lower cost to consumers. Even during extreme weather events like the 2014 Polar Vortex and Winter Storm

---

[38] Brattle's 2008 Review of PJM's Reliability Pricing Model (RPM) (PDF)

Elliott in 2022, the capacity reserves procured through the RPM helped provide the "bench strength" needed to help maintain system reliability even as fleet-wide average performance fell short of expectations.

Figure 5. RPM Reserve Margins and Clearing Prices Over Time



2. **Coal-to-Gas Transition.** In roughly a decade, between 2011 and 2023, PJM witnessed the retirement of approximately 54 GW of generation capacity (a fleet roughly the size of the New York ISO grid), as older coal units deactivated due to environmental regulations (such as the Mercury and Air Toxins Standard and effluent limitations) and economics, and were replaced by a wave of highly efficient, low-cost natural gas generation. RPM pricing signals helped facilitate this incredible transition in technology and fuel, helping to coordinate entry and exit decisions all while maintaining robust reserve margins.

3. **Demand Response.** The RPM construct was designed to enable competition among differing types of capacity, including demand response. Since its inception in 2007, the amount of demand response on the PJM system has grown significantly from being slightly above 100 MW to a level that has consistently hovered between 8,000 MW and 10,000 MW by the mid-2020s. The market has helped to drive the innovative success of the capacity product, which has provided significant reliability benefits and savings to the grid over the years, as well as helped smooth the fleet transition from coal to gas plants.

4. **Private Investment vs. Public Risk.** The RPM and PJM's competitive markets have attracted more than 40 GW of new generation over the past two decades with billions in capital invested, with much of that funding coming from private market capital rather than traditional utility rate-basing. This has all occurred under a market design that places the investment risk with private developers rather than having that risk placed on the public or captive "ratepayers" as it was under the traditional utility model.

Role of Electric Distribution Companies and States Under RPM

It is important to acknowledge that resource adequacy has traditionally been and continues to be a shared responsibility within the PJM footprint. While PJM is responsible for many aspects of the resource adequacy construct – including the assessment of capacity requirements to meet the pool-wide LOLE standard, determining zonal shares of those requirements and operating the wholesale capacity market – states and Electric Distribution

Companies (EDCs) also play a critical role, with responsibilities and/or authority to directly control or impact the following areas of resource adequacy:

- **Capacity Cost Allocation:** EDCs are responsible for determining the methodology and allocation of wholesale zonal capacity obligations to individual customers and LSEs within their service territory. These values are provided by EDCs to PJM to determine each LSE's wholesale capacity costs. Furthermore, states have authority over retail cost allocation and setting how costs are allocated to the various customer classes in utility tariffs.
- **Load Interconnection:** The requirements and rules regarding the interconnection of new loads on the system, including data centers, largely rests with the EDC and states. They have the authority to determine the entry requirements and/or financial commitments of new loads seeking to connect to the grid that ultimately impact the demand for capacity on the system and ability to reliably serve all load. EDCs are also responsible for providing the forecasted large load adjustments for PJM's load forecast.
- **Generation Siting and Permitting:** States play a pivotal role in projects ultimately getting approved and developed through their siting/permitting processes. The efficiency and effectiveness of those processes have a direct impact on the ability of new generation to come online and meet projected demand.
- **Load Shedding Priority:** Ultimately, when there is inadequate generation to serve the demand, PJM will enter emergency actions and may need to shed firm load on the system. PJM will assess the allocated share of load shed for each control zone, but within each zone, states and utilities ultimately have the authority to decide which customers or customer classes (and breakers), if any, are deemed critical and may have first right to the available capacity on the system.
- **Setting Environmental Policies:** States have the direct authority to set environmental targets and regulations within their states. These policies are one of the mechanisms that the state can use to incentivize a desired technology or portfolio of technologies within a state. Many states in PJM have chosen policy approaches that prefer renewable technologies such as solar and wind over fossil generation while others have sought to outlaw emitting technologies.

Appendix B. Foundations of PJM's Energy & Ancillary Services Markets

This appendix provides an overview of the early foundations of E&AS markets in North America to provide additional context for the E&AS market reforms discussed in the report body.

The Wholesale Power Market Platform (WPM) design proposed by FERC in 2003 represented a significant step toward the creation of a more efficient, transparent and competitive electricity market.[39] Versions of this design have been implemented in seven North American ISOs/RTOs, including PJM. The core features of this design addressed several key challenges in the operation and management of power systems and remain in place today. They included:

- A wholesale power market operating over a high-voltage transmission grid: This feature ensures that the market can facilitate the efficient exchange of power across a wide geographic area, taking advantage of the economies of scale and diversity of resources that a larger grid provides. The high-voltage transmission grid allows for the transfer of power from regions with surplus generation to regions with higher demand, promoting the optimal use of resources and reducing overall system costs.
- Central management by an independent system operator: The independent system operator (ISO) plays a crucial role in ensuring the fair and efficient operation of the market. By acting as a neutral party, the ISO can make decisions that prioritize the overall reliability and cost-effectiveness of the system, rather than the interests of any particular market participant. The ISO is responsible for managing the day-ahead and real-time markets, as well as ensuring the secure operation of the transmission grid.
- A two-settlement system with day-ahead and real-time markets: The two-settlement system was designed to promote efficient price formation and risk management. The day-ahead market allows market participants to commit to buying or selling power in advance, based on their expectations of future supply and demand conditions. This forward market helps to reduce uncertainty and promote more efficient scheduling of resources. The real-time markets, on the other hand, provide a mechanism for fine-tuning these commitments in response to actual conditions on the day of delivery. By operating these markets in tandem, the two-settlement system helps to ensure that supply and demand are continuously balanced, while also providing flexibility to adjust to changing circumstances and system conditions.
- Locational marginal pricing (LMP) for congestion management: LMP was a key innovation in the design of wholesale power markets. By pricing power according to the location and timing of its injection or withdrawal from the grid, LMP helps to reveal the true cost of delivering power to different points in the system. This is particularly important in the context of grid congestion, where transmission constraints can limit the ability to transfer power from one location to another. By assigning higher prices to locations where congestion is present, LMP creates incentives for generators to locate in areas where they are most needed, and for consumers to shift their demand to times and locations where power is more readily available. This helps to alleviate congestion and improve the overall efficiency of the system.

---

[39] FERC White Paper, "Wholesale Power Market Platform," April 28, 2003, Docket No. RM01-12-000
See also discussion introducing White Paper in the Federal Register, Vol. 68, No. 89, Thursday, May 8, 2003.

- Market power oversight and mitigation: The WPM design acknowledged the need to prevent dominant market participants from exercising market power, which can arise due to factors such as ownership concentration, transmission constraints and inelastic demand. To mitigate these risks, the design included provisions for market power oversight and mitigation, such as price caps, offer caps and bidding behavior restrictions, ensuring competitive prices, and undistorted market outcomes. Any evolution of PJM's capacity market design – whether toward mandatory hedging, centralized long-term procurement or a shift toward energy market revenue recovery – would need to adapt the IMM's mitigation role accordingly. Ensuring that market power does not taint bilateral or centralized contract outcomes is a necessary complement to any structural reform, not a detail to be resolved after the fact.

While the WPM design represented a significant step forward in the evolution of electricity markets, it was not a panacea. As in the capacity markets, this design relied on several assumptions and simplifications, as the following sections will describe. It has been, and will continue to be, essential to build on these original foundations, adapting and innovating to meet the needs of a rapidly changing grid.

Energy Market Pricing Fundamentals

In general, prices "clear" a market. The clearing price is the price at which the quantity that suppliers are willing to sell equals the quantity that buyers are willing to buy. Economic theory guarantees that under certain assumptions (including but not limited to convex cost functions, perfect competition and risk neutrality), prices alone can provide the revenue needed to support efficient operations and optimal investment over the long term. (In fact, this result largely motivated the transition from deregulated markets over two decades ago.) These prices are not only efficient but also the natural result of willing buyers and sellers transacting in a competitive environment, each trying to maximize their own welfare or profits.

PJM's value proposition is to leverage markets to incentivize least-cost reliability, which necessitates prices that reflect value so that profit-maximizing sellers can invest and operate in ways that deliver the most value. But markets that directly and transparently reflect supply and demand fundamentals and yield efficient clearing prices do not arise naturally for electricity as they do for other commodities. Electricity markets are unique due to several factors:

- The need to balance supply and demand second-by-second, with potentially catastrophic costs for failure to do so (shedding load is bad enough; cascading outages and network collapse are far worse)
- The virtual impossibility of enforcing contracts, as grid operators cannot stop consumers from taking electricity even if they haven't arranged to purchase it in advance
- The costliness of storage, resulting in low storage capacity (as opposed to cheap warehouses for most commodities) and limited price arbitrage over time (which would tend to smooth relative fluctuations in supply and demand)

While not unique to electricity markets, the non-convex (blocky) cost of supply resources adds further complexity. These characteristics make electricity markets different from most other markets, so the problems encountered and market design solutions needed look different as well.

The physical need to continually balance supply and demand requires that real-time electricity markets clear at high frequency, ensuring there is neither surplus nor shortage at any node given the pattern of injections and withdrawals respecting physical constraints on the system, such as transmission voltage and stability limitations and resource constraints like startup time, ramp rate, up and down time, and output constraints. At the same time, markets necessarily make compromises relative to the full complexity of electricity systems, such as using linear power flow assumptions and discretizing continuous time. These simplifications reflect trade-offs among accuracy, simplicity and computational feasibility. Another important simplifying assumption was that, with few exceptions, load was load and was not price responsive or dispatchable. Given changes in the resource mix, the potential for demand flexibility and significant growth in computational power, the compromises that were optimal when the energy markets were founded over two decades ago warrant reevaluation.

Appendix C. Long-Term Energy Hedging: The Spectrum of Options for Consumer Protection

A fundamental question underlying E&AS market reform is: If we allow scarcity prices to reflect the true value of reliability, how do we protect consumers – particularly those served by default service providers – from the resulting volatility? The answer lies in a spectrum of long-term forward energy hedging instruments, each representing a different allocation of risk between generators, Load Serving Entities, consumers and the system operator. Understanding this spectrum is essential for evaluating both the near-term affordability reforms available to PJM today and the longer-run design options associated with Path C.

Decentralized LSE Hedging: The Baseline Case

The simplest approach is to rely on Load Serving Entities to voluntarily hedge their load through forward energy contracts – Power Purchase Agreements (PPAs), tolling agreements and financial energy swaps – entered bilaterally with generation owners. This is already the standard practice for competitive LSEs in PJM. The problem is not that such contracts are unavailable; it is that state-regulated default service providers, operating under rate-of-return constraints and annual procurement cycles, have historically been constrained from entering long-term bilateral contracts by state commission requirements and procurement rules optimized for capturing low spot market prices during the surplus era.

Reforming the regulatory framework to permit – or require – state-regulated LSEs to execute longer-term bilateral energy hedges is the most direct and least administratively complex path toward resolving the credibility trap. It does not require PJM to administer a new centralized product; it requires state regulators to update procurement rules that were written for a different market environment.

Mandatory Forward Hedging Requirements

Stanford economist Frank Wolak has developed the most rigorous academic framework for mandatory long-term forward contracting as an alternative to capacity markets.[40] Wolak's core argument is that the political instability of capacity market price signals – the mechanism we have called the credibility trap – is not a failure of market design per se, but a consequence of the reliability externality created by offer caps. When energy price caps are set below VOLL, retailers face limited financial consequence for failing to hedge their load: the cap limits their exposure to scarcity prices, removing some of the economic incentive to contract forward. The result is systematic underhedging, which in turn makes unhedged consumers vulnerable to price spikes – triggering the intervention cycle that destabilizes investment signals.

Wolak's proposed solution – the Standardized Fixed-Price Forward Contract (SFPFC) – is itself a form of mandatory LSE procurement Requirement, but with a distinctive supplier-side incentive structure. LSEs are required to hold SFPFCs covering specified fractions of system Demand at different forward horizons. These contracts are procured through descending-price auctions administered by the market operator. On the supply side, generator owners participate voluntarily: each unit may sell up to a maximum quantity of SFPFC energy. A generator that has sold

---

[40] Frank A. Wolak, "Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables," *Environmental and Energy Policy and the Economy*, vol. 3 (University of Chicago Press, 2022), pp. 155–220. See also Wolak, "Regulating Competition in Wholesale Electricity Supply," in *Economic Regulation and Its Reform: What Have We Learned?*, ed. Nancy L. Rose (University of Chicago Press, 2014).

SFPFC energy must financially settle the difference between its contracted price and the realized hourly spot price across its full obligation, creating strong incentives to produce during those high-value hours in which buying back at system prices would be most costly.

Wolak's framework draws on empirical evidence from Chile and Peru – both of which operate "supplier-only short-term markets" in which retailers are required to purchase full-requirements contracts from generators. Chile has successfully served demand with sustained annual growth exceeding 7% since 1992 under this structure, without a U.S.-style capacity market. The analogy to the SFPFC is direct: both mechanisms place the legal purchase obligation on retailers while subjecting suppliers to the financial risk of short-term market outcomes through their voluntary forward sales positions.

For the PJM context, the feature of the SFPFC that distinguishes it from a simpler LSE bilateral hedging mandate is the supplier-side financial incentive mechanism. A regulatory requirement directed at state-regulated LSEs to maintain minimum forward hedge positions captures the consumer protection dimension of Wolak's argument – insulate consumers from spot volatility, remove the political incentive for intervention – without replicating the full SFPFC structure. What the SFPFC adds is a formalized auction mechanism and standardized contractual form for the product that creates a direct private incentive to ensure energy availability every hour of the year, and especially in the highest price hours.

The key insight shared by both formulations is that mandatory forward hedging requirements do not require the hedging to be done in a capacity product. A Load Serving Entity that has entered a 10-year PPA at $45/MWh for 80% of its expected peak load is, for practical purposes, financially insulated from a $500/MWh real-time energy price – whether or not it has also procured the corresponding capacity obligation. The energy contract provides the consumer protection that the capacity market's hedging function was intended to deliver in a more standardized and more project-finance-legible instrument.

Centralized Procurement of Reliability Options

If the region determines that relying on decentralized or mandatory LSE hedging leaves unacceptable residual risk of consumer exposure during scarcity events, PJM could administer a centralized hedge through a product structure known as the Reliability Option, already deployed in several international jurisdictions.

Peter Cramton and Steven Stoft developed the theoretical framework for reliability options, arguing that the right instrument for bridging energy-only market design and consumer protection is a financial derivative on scarcity prices, not an administrative procurement of "installed capacity" measured in megawatts.[41] A Reliability Option is structured as a physical capacity commitment bundled with a financial call option: The generator sells the option in a centralized auction, receiving a premium that supports project financing, and commits to rebate above-strike spot market revenues to load.

The mechanism operates through a three-party settlement. When the real-time price exceeds the strike, the generator owes load a payment equal to the strike price deviation (spot price minus strike price) multiplied by the generator's contracted quantity share – not its actual real-time output. This design is crucial: because the generator's financial obligation is proportional to its contracted share rather than its actual dispatch; it enters the spot market with a nearly balanced position during scarcity. Market power is substantially reduced precisely when it is most

---

[41] Peter Cramton and Steven Stoft, "Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency," *Utilities Policy* 16(3) (2008), pp. 194–201. The Reliability Option concept is also developed in Cramton and Stoft, "The Convergence of Market Designs for Adequate Generating Capacity," White Paper for the Electricity Oversight Board (2006).

problematic – during high-price, low-supply events – because generators that attempted to withhold output to drive prices higher would bear the above-strike rebate on their contracted share regardless of whether they dispatched.

One important design consideration for any Reliability Option mechanism is the cost of the option premium itself. Long-dated options on illiquid underlyings are expensive: Standard derivatives pricing (Black-Scholes) relies on risk-neutral replication, which is not feasible when the underlying asset cannot be liquidly traded. In practice, this means the premium includes not just a hedging cost but an "insurance-style" risk premium for the unhedgeable residual. This is a genuine cost that policymakers should not underestimate. The mechanism used in practice – competitive centralized auctions, as in the UK and Irish/Northern Irish markets – addresses this by allowing the market to determine the premium price rather than deriving it analytically. Competitive auction pricing incorporates all participants' views on cost and risk and has demonstrated commercial bankability in the UK and Ireland contexts. But the fundamental point stands: Centralizing and standardizing the option does not eliminate the underlying option cost; it makes it transparent and competitively determined rather than embedded and opaque.

The key terms of a Reliability Option:

- **The premium**: PJM, acting on behalf of LSEs, pays generators an up-front premium in exchange for their commitment. This provides stable, bankable revenue that supports project financing – serving the same function as the current capacity payment but embedded within an instrument that is directly integrated with E&AS market reform.
- **The strike price**: Set at or near VOLL to preserve optimal dispatch incentives. During normal market conditions, the generator retains all energy market revenues. During scarcity events, the generator's above-strike revenues are returned to load.
- **The performance incentive**: If a generator fails to deliver energy during an event when the price exceeds the strike, it still owes the rebate – creating a strong market-based performance incentive without a separate administrative penalty structure.
- **Contract terms**: Cramton and Stoft propose that new generation receives contracts of up to seven years, providing the revenue certainty needed for construction financing; existing generation receives one-year contracts consistent with their lower capital recovery needs.

The Reliability Option mechanism resolves the central tension in E&AS market reform: scarcity pricing that is economically necessary but politically unsustainable. Under a Reliability Option regime, the spot market is free to clear at $5,000/MWh or $15,000/MWh during a grid emergency – providing the full economic signal to dispatch flexible loads, incentivize generator performance and attract investment. But consumers see their effective energy cost capped at the strike price because the above-strike revenues are returned to load through the financial rebate. The market signal and the consumer protection operate simultaneously in different instruments.

Centralized Forward Energy Market

A more recent proposal from Cramton and coauthors takes a different architectural approach to the problem of forward market illiquidity.[42] Rather than a Reliability Option administered through a periodic big-event auction, their 2025 design proposes a centralized forward energy market, operated by the system operator, in which market participants trade thousands of granular forward contracts and European call options continuously, up to four years ahead.

The core design elements are:

- **Granular products**: Participants can trade forward contracts for every combination of year-month-hour-weekend/weekday for up to 48 months ahead, plus daily products for the next 30 days. This granularity allows generators and LSEs to build hedge positions that closely match their actual seasonal and diurnal load and generation profiles. Existing OTC markets offer only broad monthly, quarterly or annual products, which cannot hedge the shape mismatches that drive residual exposure.
- **Flow trading technology**: The market applies the "flow trading" methodology (Budish et al., 2023) to clear thousands of products simultaneously. Participants submit persistent portfolio orders expressing net demand curves; hourly batch auctions clear all products at once via a quadratic programming routine. This makes gradual, small-quantity trading over time feasible at scale, mitigating adverse price impact and reducing market power opportunities. The system operator publishes aggregate net demand slopes as liquidity indicators, making the market highly transparent.
- **High-strike call options**: In addition to forward contracts, the market offers European call options with a strike price set at a level like $1,000/MWh. The high strike means the option is relevant only for true tail scarcity events; it supplements the forward position without substituting for it. The paper explicitly distinguishes these from the physical-capacity-bundled Reliability Options of Cramton-Stoft (2008): These are purely financial instruments that allow LSEs to hedge the residual tail risk that forward energy contracts leave unhedged.
- **Mandatory LSE purchase obligation**: To bootstrap liquidity and coordinate trade, the design includes a regulatory obligation for LSEs to purchase forward cover. The obligation ramps linearly from zero at 48 months ahead to 100% of realized real-time load one day ahead. This linear schedule differs from Wolak's steeper ramp (85% at four years ahead) because gradual accumulation allows LSEs to incorporate new information as it arrives, reducing price impact and making risk management more tractable.
- **System operator role**: Critically, the authors argue that this market *should be* operated by the independent system operator rather than private exchanges. Private exchanges lack both the incentive and the product structure to provide adequate liquidity for products more than one year ahead; their revenue models depend on charging for services – data feeds, co-location – that are byproducts of a flawed trading format. The system operator alone is motivated to adopt an efficient and transparent forward market.

---

[42] Peter Cramton, Simon Brandkamp, Jason Dark, Darrell Hoy, David Malec, Axel Ockenfels and Chris Wilkens, "A Forward Energy Market to Improve Reliability and Resiliency," University of Maryland Working Paper (March 2025). The paper extends a 2024 policy white paper by the same authors. The flow trading methodology it applies is developed in Eric Budish, Peter Cramton, Albert S. Kyle, Jeongmin Lee and David Malec, "Flow Trading," NBER Working Paper No. 31098 (2023).

The paper develops a twelve-year simulation of ERCOT's day-ahead market (2011–2022) as a proof of concept. A key finding is that agents' risk preferences critically shape their demand for forwards and options: At moderate risk aversion, forward market clearing prices converge to within 10% of arbitrage-free benchmarks. The design explicitly contemplates an incremental implementation path – beginning with a 30-day forward market while retaining existing capacity requirements, then extending the trading horizon as stakeholders gain familiarity with the mechanism.

The policy implication the authors draw is direct: a well-functioning centralized forward energy market would allow the ISO to raise the real-time price cap (because participants are in near-balanced positions and sufficiently hedged), address nonconvex cost recovery through forward price signals, and potentially replace or substantially reduce the relevance of capacity markets. The missing money problem does not disappear – but it becomes manageable through forward energy market design rather than an administratively separate capacity product.[43]

For PJM, this design is further out on the implementation spectrum than voluntary bilateral hedging or mandatory LSE requirements – it requires the ISO to operate a new centralized market infrastructure. But it addresses a genuine gap that the other mechanisms do not: Bilateral OTC markets simply cannot provide the granular, multiyear forward price signals that investment decisions require, and private exchanges have not solved the liquidity problem for tenors beyond one year. The Cramton et al. design represents a market-structure answer to that gap.

International Experience: Ireland and Australia

The Reliability Option framework and related centralized long-term energy contracting approaches have been implemented or explored in other jurisdictions.

The Irish and Northern Irish Single Electricity Market (SEM) operates a capacity remuneration mechanism with explicit reliability option structure – contracting around a derivative of the energy product. Units clearing in the auction hold delivery obligations during high-stress periods, with performance requirements enforced throughout the delivery year.[44]

In Australia, the National Electricity Market has operated as an energy-only market for three decades, with a high administered price cap (currently AUD $16,600/MWh) providing the scarcity pricing function. As the NEM transitions away from dispatchable coal generation and reliability risk increases, the Australian government launched the Capacity Investment Scheme (CIS) in 2023 – not a U.S.-style capacity market, but a centralized procurement of long-term revenue floor/ceiling agreements for new generation and storage at 10- to 15-year terms. The CIS offers two contract structures: a Clean Dispatchable CISA (for dispatchable capacity with at least two hours of storage) and a Generation CISA (for variable renewable generation). Both structures operate with a revenue floor – the government pays the project when market revenues fall below the contracted floor – and a revenue ceiling – the project returns an agreed share of revenues to the government when revenues exceed the

---

[43] Peter Cramton, Simon Brandkamp, Jason Dark, Darrell Hoy, David Malec, Axel Ockenfels and Chris Wilkens, "A Forward Energy Market to Improve Reliability and Resiliency," University of Maryland Working Paper (March 2025). The paper extends a 2024 policy white paper by the same authors. The flow trading methodology it applies is developed in Eric Budish, Peter Cramton, Albert S. Kyle, Jeongmin Lee and David Malec, "Flow Trading," NBER Working Paper No. 31098 (2023).

[44] SEM (Ireland/Northern Ireland) T-4 auction data: EirGrid/SONI, *T-4 Capacity Market Auction Overview: Delivery Year 2025/26* (2022).

ceiling. The ceiling preserves the project's exposure to high energy prices, maintaining dispatch incentives in the spot market while the floor provides the financing certainty needed to reach investment decisions.[45]

The most current thinking on Australian market design comes from the December 2025 NEM Wholesale Market Settings Review, an independent review commissioned by the Australian Department of Climate Change, Energy, the Environment and Water.[46] The review panel concluded that Australia should retain its real-time energy-only spot market as the core market structure – recommending explicitly against introducing a U.S.-style capacity market. But the panel identified a critical gap: the "tenor mismatch" between the long financing timelines required by infrastructure investors (typically 7–15+ years) and the short contracting horizons that market buyers can currently offer (typically 0–3 years). Forward contract markets cannot bridge this gap organically, leaving new generation projects exposed to unhedged revenue risk beyond their near-term contract periods.

To address this gap, the review recommended a new statutory mechanism – the Electricity Services Entry Mechanism (ESEM) – to be embedded in Australian National Energy Law and Rules. ESEM would procure standardized, fungible long-term contracts for three categories of grid services: bulk energy, shaping (moving supply toward high-demand periods) and firming (providing dispatchable backup during low-supply events). These contracts would cover the "later years" of a project's commercial life (years 3–15+), where neither bilateral buyers nor derivative markets can currently offer hedging. The contracts would be standardized rather than bespoke, making them fungible instruments that can be traded between market participants as conditions change – creating an ESEM secondary market that did not exist under the CIS government underwriting model. The review also recommended a market making obligation (MMO) to ensure liquidity in the existing derivative markets that serve the shorter-tenor hedging needs.

The ESEM represents a significant evolution beyond the CIS model. Where CIS was government underwriting for specific projects, ESEM is a structured market mechanism: It procures standardized instruments, recycles them back to market participants who need hedging, and operates within a regulated market framework rather than as a discretionary government program. This distinction is directly relevant to PJM's situation: The ESEM model proposes that it is possible to address the tenor gap – the same mismatch between investor financing horizons and buyer contracting horizons that characterizes the PJM E&AS market – without either a U.S.-style capacity market or open-ended government subsidy.

---

[45] Australian Government, Department of Climate Change, Energy, the Environment and Water, *Capacity Investment Scheme: Design Paper – Western Australia Implementation* (April 2024). The CIS operates alongside Western Australia's existing Reserve Capacity Mechanism and is designed to complement rather than replace the existing capacity credit market structure.

[46] Tim Nelson, Paula Conboy, Ava Hancock and Phil Hirschhorn, *National Electricity Market Wholesale Market Settings Review: Final Report* (Australian Government, Department of Climate Change, Energy, the Environment and Water, December 2025), Recommendations 1, 4–7 (ESEM design), and supporting analysis at pp. 165–175

Comparative Assessment and Relevance to PJM

The hedging instruments described above span a spectrum from market-led to administratively intensive:

Table 2. Summary of Alternative Long-Term Energy Hedging Options

Approach	Revenue Instrument	Administrative Intensity	Timeline
Voluntary bilateral PPAs	Energy forward contract	**Minimal** (market-led)	Immediately available
Mandatory LSE hedging requirements	Energy forward contracts (required)	**Moderate** (regulatory)	1- to 2-year rulemaking
Mandatory generator forward contracts (SFPFC)	Fixed-price full-requirements contracts	**Moderate–High** (regulatory + market rules)	2- to 3-year implementation
Centralized Reliability Option	Centrally auctioned financial option	**High** (PJM-administered)	3- to 5-year implementation
Centralized forward energy market (ISO-operated)	Granular hourly forward contracts + call options; mandatory LSE purchase ramp	**High** (ISO market infrastructure)	3- to 5-year implementation
Centralized long-term energy service contracts (ESEM)	Standardized fungible energy/ shaping/firming contracts (statutory)	**High** requires legal/tariff authority	In development (Australia, 2026+)

For PJM's near-term situation, the most actionable interventions are at the left end of this spectrum: regulatory reforms to permit and incentivize state-regulated LSEs to execute longer-term bilateral energy contracts, combined with mandatory hedging requirement frameworks. These reforms are compatible with both Path A (as a stabilization measure that insulates load from spot price volatility while the capacity market continues to function) and Path C (as the first step in a longer-run transition toward an energy-market-centric design).

The centralized options – Reliability Options, ISO-operated forward energy markets and ESEM-style instruments – are more administratively intensive and have longer implementation horizons but represent logical endpoints for a region that determines bilateral hedging leaves residual exposure requiring centralized protection. The Cramton-Stoft (2008) framework provides the theoretical foundation for a centralized financial option; the Cramton et al. (2025) design offers a more comprehensive answer to the forward market liquidity problem itself, addressing the missing money problem through ISO-operated granular markets rather than through a separate capacity product; and the Australian ESEM demonstrates that the tenor gap at the long end of the financing horizon can be addressed through standardized statutory instruments. The international evidence collectively demonstrates that centralized energy contracting mechanisms are workable at scale.

Ultimately, evolving the E&AS markets to reflect true scarcity does not condemn consumers to unaffordable bills. Rather, it allows the market to use price signals to efficiently manage grid operations while financial hedging instruments – at whatever point on the spectrum is appropriate to the region's political and regulatory context – manage consumer cost exposure. The right combination of E&AS market reform and long-term hedging design is the necessary architectural foundation for any of the three paths forward described in this paper.