# Unbundling the Reliability Compact

**Christoph Graf**[1]
(Program on Energy and Sustainable Development, Stanford University)

**Joshua C. Macey**
(Yale Law School)

**Jacob Mays**
(Civil and Environmental Engineering, Cornell University)

## Introduction

Forward capacity markets emerged in the mid-2000s as a response to the problem of ensuring adequate investment in newly liberalized power markets. The conventional case for capacity markets arose from the interaction of three features: (1) most retail consumers did not face real-time prices and therefore lacked incentives to reduce consumption when electricity became scarce; (2) scarcity prices were frequently capped or otherwise suppressed, muting investment incentives and preventing energy-market revenues from supporting enough investment to meet peak demand; and (3) system operators could not readily curtail customers with the lowest willingness to pay for reliability. Centralized capacity markets were designed to address these issues by procuring an administratively determined quantity of capacity in advance and allocating its cost to LSEs.

In the two decades since PJM introduced the Reliability Pricing Model (RPM), it has become evident that forward capacity markets suffer from difficult and perhaps insoluble problems. One challenge is that RPM is administratively complex. It requires PJM to forecast uncertain future demand, translate heterogeneous resources into a common capacity metric, determine regional and locational procurement requirements, select a reference technology and estimate its cost of new entry, construct an administrative demand curve, and define and enforce resource performance obligations. These tasks have only become more difficult as the resource mix shifts toward storage, variable generation, demand response, and fuel-constrained resources. Most recently, rapid and lumpy growth from new large loads has added a further forecasting problem.

---

[1] Christoph Graf is also affiliated with the Institute for Policy Integrity at NYU School of Law. The views expressed in this paper are solely those of the author and do not represent the views or positions of the institutions with which the author is affiliated.

Largely in response to the concerns that RPM is not functioning as intended, PJM recently issued a whitepaper proposing three potential paths for reforming its resource adequacy mechanism. Path A would retain RPM as the principal resource adequacy mechanism while requiring load-serving entities to secure longer-term capacity commitments, or alternatively, authorizing PJM to conduct centralized long-term procurement. Path B would permit differentiated reliability arrangements under which some customers—particularly new large loads—could connect more quickly or at lower cost in exchange for accepting enforceable curtailment risk. Path C would shift a larger share of resource revenue recovery toward the energy and ancillary services markets, protect customers through forward energy contracting, and retain a reduced capacity mechanism as a backstop. Path B could in theory complement either Path A or Path C; Paths A and C represent different choices about the central pricing and contracting architecture.

These paths can be evaluated by separating four analytically distinct functions that are linked together in PJM's current reliability architecture. The first is the **reliability standard**: the metric, risk measure, and target used to define an acceptable level of involuntary curtailment of firm load. The second is **service entitlement**: which demand is treated as firm, whether some customers may voluntarily accept lower-priority service, and how shortages are allocated when all demand cannot be served. The third is **operational coordination and pricing**: how resources are committed and dispatched, how reserves are procured, and how realized scarcity and flexibility are reflected in market prices. The fourth is **financial risk allocation**: how price, volume, performance, and default risk are distributed through forward contracts, collateral requirements, and retail arrangements. Taken together, choices across these four functions shape the system's **physical adequacy and investment efficiency**: whether a deliverable portfolio of generation, storage, transmission, and demand-side capability is built and retained at reasonable cost to meet the selected reliability standard.

Viewed through this framework, Path B is primarily a reform to service entitlement, although it also affects operations and the quantity of resources needed for physical adequacy. Paths A and C differ principally in where performance incentives appear and what underlying exposure is used for long-term contracting. Path A would continue to use an administratively defined capacity product both to support physical adequacy and as the reference instrument for long-term hedging. Path C would instead place greater weight on time- and location-specific energy and ancillary services outcomes as the reference for operational performance and forward risk allocation.

Our primary recommendation is that PJM pursue Path C as its primary approach to resource adequacy, complemented by demand-side reforms such as those proposed as part of Path B.

Both Paths A and C better allocate financial and performance risk by putting more responsibility on LSEs and other customers, but Path C contains a number of advantages that arise because it more efficiently unbundles the different functions and objectives currently bundled in centralized capacity markets. Specifically, it allows LSEs to assemble portfolios that reflect their customers' load profiles, clean energy obligations, and price exposure, while permitting suppliers to use contract forms that suit their technologies. Path C also reduces the extent to which investment depends on changing capacity market methodologies and assumptions. And, because most load would be hedged in advance, scarcity pricing would be more credible because retail customers would not be immediately exposed to the full cost of those prices.

Under this approach, improved energy and ancillary services price formation would provide the principal incentives for realized availability and operational performance. Forward energy contracting would protect consumers from volatility, support investment, and accommodate contract forms suited to different types of resources. Customers could contract for firm or interruptible service depending on their needs and willingness to pay for reliability.

# PJM's Three Paths

PJM recently proposed three strategic pathways to reform its resource adequacy mechanism with the goal of creating a more durable framework for consumers. Path A would reform the existing capacity market. Path C would shift towards a model that emphasizes scarcity pricing. Path B would move away from the shared reliability compact by allowing resources that have a lower willingness to pay for reliability to accept curtailment risk in exchange for expedited interconnection service and lower resource adequacy payments.

**Path A (Preserving the shared reliability compact and long-term capacity hedging):** Path A would maintain a uniform standard of grid reliability where all customers share the costs and physical benefits of resource adequacy. The capacity market would continue to exist, and capacity market spot prices could be used to trade short-term capacity deviations and act as underlying prices for long-term contracting. High capacity market spot prices would signal scarcity and provide incentives to invest. However, long-term capacity hedging (long-term commitments) would protect consumers from sudden rate shocks. PJM outlines two potential procurement structures:

- **LSE Hedging Requirements:** Load Serving Entities (LSEs) must submit resource plans demonstrating they have hedged a high threshold (e.g., at least 90%) of expected obligations years in advance through physical assets or bilateral contracts, leaving the spot capacity market to function purely as a residual market for short-term deviations.

- **Centralized Long-Term Procurement:** PJM centrally administers a long-term, laddered procurement of capacity (e.g., 7-year contracts covering up to 70% of obligations) and directly determines contract lengths and maximum willingness to pay. This pathway stabilizes consumer costs and provides developers with greater revenue certainty to finance new generation and storage resources.

**Path B (Differentiated reliability allocations):** This approach abandons the assumption of a universal reliability standard, accepting that reliability must be explicitly rationed during structural scarcity. Rather than managing deficits via collective load-shedding, PJM would differentiate reliability based on class, geography, or cost causation. Customers would select service contracts (similar to cellular plans) that determine their specific pricing and curtailment priority during emergencies. A primary feature is the "Connect and Manage" (or Interim Resource Adequacy Service (IRAS)) framework: large new loads (such as hyperscale data centers) that connect before sufficient additional deliverable supply or demand-side capability is available would be placed ahead of existing firm load in the curtailment queue. The goal of that framework is to prevent new entry from increasing the outage risk borne by other customers.

**Path C (Emphasize energy market revenues and long-term energy hedging):** Consistent with Schweppe's real-time pricing framework, this model treats delivered energy as the fundamental commodity and shifts the primary revenue recovery function from administrative capacity markets to the daily Energy and Ancillary Services (E&AS) markets. E&AS price limits are raised to allow energy prices to reflect true scarcity (potentially spiking to $10,000/MWh or higher during periods of scarcity). To address credit risk and shield consumers from extreme real-time price volatility, LSEs would be required to hedge. The capacity market under this framework would be retained as a residual backstop, but its importance would diminish as suppliers' revenues would be determined primarily by contractual obligations tied to the E&AS market, with any differences settled through energy market prices.

# Assessment of PJM's Three Paths

In this section, we assess the implications of the three paths outlined in PJM's proposal across four dimensions: the choice of reliability objective; service entitlements; operational coordination and pricing; and financial risk allocation.

## Reliability Standard

Before choosing a long-term resource adequacy mechanism, policymakers must determine the reliability standard the mechanism is intended to achieve. Resource adequacy is not an ordinary commodity whose quantity and price can be observed directly. It is a probabilistic planning construct based on simulations of demand, resource availability, weather, and system operations across possible future conditions. PJM's current standard is commonly summarized as the "one-day-in-ten-years" criterion. That criterion limits the planned expected frequency of days with a loss-of-load event, but is largely agnostic to the magnitude, duration, and geographic extent of the resulting shortfalls. Other metrics, such as expected unserved energy, can reflect the magnitude and duration of shortfalls, while tail-risk measures can place greater weight on unusually severe outcomes. These choices are not neutral with respect to investment outcomes: different combinations of metrics, risk measures, and targets express different judgments about the types of outage risk the region should prioritize, leading to different accreditation values and therefore capacity market outcomes.

The calculated level of reliability also depends on a number of modeling choices. For example, a variety of specifications, including scenario construction, weather years, load forecasts, correlated outages, fuel availability, transmission constraints, and the modeled behavior of storage and flexible demand, can all affect the estimated likelihood and severity of shortfalls. A reliability standard is therefore best understood as a planning benchmark, not a guarantee that shortages exceeding the reliability target will never occur. Even a well-designed procurement mechanism can fail to achieve the intended standard if its underlying models omit important risks, or if the future differs materially from the scenarios considered.

Administrative choices that influence the reliability standard have to be made under all three of PJM's pathways. For example, under Path A, after the grid operator selects a reliability standard, it must impose regional and locational capacity requirements, calculate accreditation values, and make a capacity demand specification. Similarly, Path C would continue to require PJM to assess prospective reliability, both to determine whether market-supported investment appears sufficient and to establish the size of any residual capacity backstop. Under Path B, the grid operator must prospectively identify load that can be curtailed during scarcity and determine the amount of firm demand against which the standard is applied.

None of the pathways eliminates the need for a collective judgment about acceptable outage risk. PJM has and will continue to have a central technical role in producing the probabilistic analysis, but the choice of standard is ultimately a public policy determination with substantial cost and distributional consequences. The relevant comparison among Paths A, B, and C instead lies in how each translates that judgment into service entitlements, operational incentives, contracts, and investment.

## Service Entitlement

We now turn to the question of which customers are entitled to the service the reliability standard seeks to meet. Centralized capacity markets generally treat resource adequacy as a common and homogeneous service, with most customers as firm loads that are entitled to broadly similar levels of reliability. Customers can and do obtain greater reliability through backup generators, batteries, and other private investments, but cannot ordinarily gain priority to scarce grid-supplied electricity or reduce the capacity obligation assigned to its LSE unless they are incorporated in an enforceable curtailment agreement.

One reason grid operators adopted this approach was that economic and network constraints made it difficult for LSEs to procure differentiated reliability service. Historically, retail customers paid fixed rates and therefore had little reason to reduce consumption when wholesale prices rose. System operators also could not readily observe customers' heterogeneous willingness to accept interruption or selectively curtail customers with lower-valued uses of electricity. This was partly due to physical constraints. For example, because customers often shared distribution feeders, emergency load shedding occurred in relatively large blocks and could not reliably distinguish customers that had purchased firm service from those willing to accept interruption. An LSE or retail customer that underprocured could therefore continue to benefit from resources procured for the broader system. In this institutional setting, reliability was difficult to exclude at the individual customer level and was consequently provided as a common service.

Advanced metering, automated controls, and the operational capabilities of some large loads make it increasingly feasible to distinguish firm from curtailable demand. At the same time, Path B does not necessarily imply a lower reliability standard for customers retaining firm service. It can instead permit some customers voluntarily to accept a different service entitlement while preserving the existing standard for the remaining firm load. In other words, the reliability standard determines the acceptable aggregate risk of involuntary curtailment for a defined class of firm load, while the service entitlement framework determines which customers belong to that class and how any realized shortage is allocated.

We should note that differentiated service need not imply that electricity is rationed solely according to financial willingness to pay. Essential services, medically vulnerable customers, and ordinary households may warrant a protected reliability floor for legal, distributional, and public policy reasons. The clearest initial application of Path B is voluntary differentiated service for sophisticated customers capable of understanding, managing, and contractually accepting curtailment risk.

Path B can complement either Path A or Path C. Under Path A, verified curtailable demand could reduce an LSE's capacity obligation. PJM would still need to determine how much reduction to recognize, verify the customer's performance, and incorporate the arrangement into its load forecast and capacity procurement. Path A can therefore accommodate differentiated service, but it translates that service back into the capacity accreditation and compliance framework. Since the lack of ability to enforce targeted curtailment was one of the original motivations for capacity markets, however, it is worth noting that the technical viability of Path B itself weakens the case for Path A. Under Path C, a customer could instead reduce consumption in response to high energy and ancillary services prices or enter a contract authorizing curtailment under specified conditions. This makes Path B more straightforwardly compatible with an architecture centered on spot market outcomes. If all customers could submit accurate demand bids and prices were permitted to reflect the full value of scarcity, voluntary demand reduction could largely determine which consumption is forgone.

## Operational Coordination and Pricing

Under all three pathways, PJM would continue to operate the day-ahead and real-time energy and ancillary services markets. Centralized operation remains necessary because PJM possesses the system-wide information needed to commit and dispatch resources, maintain supply-demand balance, procure reserves, and respect transmission and security constraints. However, the different pathways would affect where the financial incentives associated with scarcity and performance appear, which resources and customers face those incentives, and how flexible demand participates in real-time operations.

Under Path A, the current division of responsibility would remain largely intact. Energy and ancillary services prices would guide ordinary commitment and dispatch, while RPM would continue to provide supplemental investment revenue and an additional performance incentive during designated scarcity intervals. While in principle the combination of energy prices and capacity nonperformance charges can create strong incentives for performance, the strength of the supplemental signal depends on administratively defined performance intervals, penalty rates, stop-loss provisions, and credit requirements. The total incentive to perform can therefore differ from the spot price, especially for non-capacity resources and flexible loads. Under Path C, a larger share of the scarcity performance incentive would appear directly in energy and ancillary services prices. This would make the same time- and location-specific signal visible to generators, storage resources, demand response, imports, and load.

Path B is different in kind because it principally changes the demand side of the short-term market. A customer accepting curtailable service could participate through a price-responsive demand bid or through a contractual obligation that authorizes PJM or the relevant LSE to curtail it under specified system conditions. Where customers can express their willingness to consume through bids and prices are allowed to rise sufficiently, the market can select voluntary load reductions before resorting to involuntary curtailment. Where such bidding is impractical or prices remain capped, service contracts and administrative curtailment priorities remain necessary. Path B can therefore complement either Path A or Path C.

Whether the incentive appears in spot prices or in a separate capacity settlement matters for physical operations as well as revenue recovery. For a conventional generator with marginal costs well below the price cap, weak scarcity pricing may have little effect on whether it produces in a particular interval. Storage, flexible demand, fuel-constrained generators, and other intertemporal resources instead make decisions based partly on expectations of future prices. Suppressing expected scarcity prices can reduce the opportunity cost of consuming fuel or stored energy before the system reaches its tightest condition. A capacity payment may replace some missing revenue, but unless its performance provisions provide an equally strong and timely marginal incentive, it does not necessarily restore the operating behavior that stronger energy and ancillary services prices would have induced.

Improvements in energy and ancillary services price formation are consequently valuable under all three pathways. Under Path A, stronger short-term prices can reduce missing money and the amount of revenue that must be recovered through RPM. Under Path B, they can support voluntary demand response and transparent settlement of load reductions. Under Path C, they are a necessary foundation of the architecture.

## Financial Risk Allocation

Both Paths A and C recognize that stronger scarcity incentives will be difficult to finance and politically sustain unless volatile short-term outcomes can be converted into more stable customer costs and supplier revenues through forward contracts. The principal difference is what exposure is hedged and who chooses the contracting portfolio. Path A would organize long-term risk sharing around an administratively defined capacity obligation and the residual RPM price. Path C would organize it primarily around prices for delivered energy and ancillary services.

There are several reasons to facilitate or require some degree of forward contracting, and both proposed approaches have benefits compared to PJM's existing forward capacity market. Most importantly, direct exposure of retail customers or default-service providers to extreme scarcity prices is unlikely to be politically durable. Contracts can shift that exposure to suppliers and intermediaries better positioned to manage it, allowing the wholesale market to form stronger prices without producing equally volatile retail bills. Forward positions can also reduce a supplier's incentive to raise spot prices, although market power may instead emerge in the forward market. Finally, barriers to long-term risk sharing can prevent otherwise economic projects from obtaining finance. Contracting can address these "missing markets" even where expected spot-market revenues would, in principle, be sufficient to support investment.

Properly designed contracts can provide this protection without suppressing marginal operating incentives. An LSE with a fixed-volume energy hedge still purchases its actual consumption at the spot price and financially settles the contracted quantity. Reducing consumption when the spot price is high therefore reduces its spot-market cost without changing the contract settlement. The seller likewise receives more stable revenue on its contracted position while remaining exposed to spot prices for deviations. Long-term contracting can thus manage financial volatility while preserving the time- and location-specific incentives created by the energy and ancillary-services markets.

The traditional argument in the "energy-only versus capacity" debate was that energy prices are too volatile to support finance and that a capacity product provides a stable hedge. Following this argument, Path A offers the apparent advantage of building on a relatively bankable source of supplier revenue. Longer commitments could improve revenue certainty along one dimension. Recent experience, however, suggests a more qualified comparison. A long-term capacity contract remains a claim on an administratively defined product whose quantity, qualification, and performance requirements can change.

Accreditation is a central source of this instability. The appropriate capacity credit changes with the resource mix and depends on methodological choices about scenario construction, outage correlations, weather years, and dispatch assumptions. Because these choices materially affect the quantity of capacity each resource may sell, methodological revisions create repeated opportunities for stakeholder advocacy. Such participation is not necessarily improper, but it makes the definition and value of the capacity product inherently contestable. Long-term capacity positions also remain exposed to changes in net CONE, the capacity demand curve, performance-assessment rules, nonperformance charges, stop-loss provisions, deliverability requirements, and the allocation of load obligations. By itself, extending the contract term does not eliminate these risks.

These recurring disputes contribute to the credibility problem that Path A is intended to address. When capacity prices appear too low to support investment, suppliers seek changes that increase or stabilize revenues. When prices rise sharply, LSEs, customers, and public officials seek caps or other protections. Individual interventions may be defensible, but repeated changes weaken confidence that future capacity revenues will remain available on the terms anticipated when an investment is made. Path A can reduce exposure to annual auction volatility, but it does not by itself eliminate the risk that administrative changes will alter the value of the underlying product.

Contractual form also matters. A capacity obligation has an option-like payoff that is comparatively well matched to a peaking resource whose value is concentrated in a small number of high-price hours. Baseload, variable, energy-limited, and demand-side resources naturally face different risks and may prefer different arrangements. Wind and solar resources often seek contracts linked to their production profile; storage may prefer a tolling or availability agreement that accounts for charging costs and intertemporal opportunity costs; and demand response may require terms governing notice, duration, and frequency of activation. Requiring compliance through a common capacity product can force resources to assume risks that are poorly aligned with what they can physically deliver and can crowd out complementary forms of risk sharing.

Path C would instead permit a broader menu of contracts settled against a common set of energy and ancillary services prices. Contracts could be entered at multiple horizons, with coverage increasing as the delivery period approaches and information improves. Given that a single standardized energy contract is not universally superior for all market participants, the advantage of the common spot market reference is that it allows LSEs and suppliers to select combinations of contracts whose price, volume, shape, and performance characteristics better match their underlying exposures.

The same flexibility applies to physical portfolios. An LSE, supplier, or intermediary could combine resources with complementary output profiles, pair variable generation with storage or firming supply, secure fuel arrangements, or use behind-the-meter resources to reduce spot exposure. LSEs could consider expected load shape, state clean energy obligations, customer service choices, and price risk together rather than procuring those attributes through separate and potentially inconsistent processes. Path A’s LSE-hedging variant would permit some portfolio choice, but all positions would ultimately need to be translated into accredited capacity. Centralized long-term procurement by PJM would place even more of the forecast and contract selection decision at the regional level.

PJM has a strong comparative advantage in operating the short-term market because it possesses system-wide information and centrally coordinates commitment, dispatch, reserves, and transmission constraints. It does not necessarily have the same advantage in selecting long-term contract portfolios for particular customers. LSEs, utilities, and state-designated procurement entities may possess better information about expected customer demand, retail service obligations, and state policies. They can also be made accountable for the costs of overcontracting, undercontracting, or selecting an inappropriate portfolio, whereas PJM does not directly bear those financial consequences.

# Consequences for Reliability and Efficiency

Having discussed the four functions, we now return to the question of whether the institutions described above will produce a portfolio capable of satisfying the planner's chosen reliability standard. In principle, either Path A or Path C can provide sufficient revenue to support investment, and forward contracts can make that investment financeable. However, we have argued that Path C has important advantages with respect to three of the four functions, making it more likely than Path A to support reliability and efficiency in a durable way.

Path A provides the clearest ex ante link between PJM's reliability standard and the resources procured to satisfy it. PJM would continue to forecast firm load, translate the reliability standard into regional and locational capacity requirements, accredit each resource's contribution, and require sufficient long-term commitments to meet the resulting obligation. At the same time, Path C does not necessitate that regulators or grid operators abandon a reliability target. In fact, under Path C, PJM would continue to determine a reliability target and evaluate whether the region is capable of meeting that target. The difference is that, instead of being translated into a reserve margin and a capacity target, the analysis would inform the calibration of scarcity prices, the design of forward contracting requirements, and the size of any residual backstop procurement. Australia's Retailer Reliability Obligation, under which the Australian Energy Market Operator assesses the relative firmness of different contracts in a manner similar to ELCC calculations, can provide an institutional analogue for this procedure.

Path C may provide a number of benefits beyond those discussed in this paper. For example, Path C may make it easier to provide connect-and-manage interconnection service and facilitate long-term transmission planning. A core challenge in the present regime, in which a significant share of transmission investment occurs through the interconnection process through participant-funded upgrades, is that projects are selected to minimize the cost to facilitate a narrow group of projects rather than to maximize the benefits to grid users writ large. High capacity prices reduce the incentive resources have to opt into energy-only interconnection service. That, in turn, means more resources request a deliverability study, increasing the administrative burden of managing the interconnection queue and reinforcing this suboptimal planning regime.

These conclusions are provisional, however, and we hope that other commentators consider the relationship between market design and transmission investment in more detail. Our goal in this paper is not to provide a comprehensive assessment of the relative advantages and disadvantages of capacity markets. Instead, we have tried to draw attention to the fact that capacity markets link four functions–a reliability standard, service entitlement, allocation of financial risk, and operational coordination and pricing–within the same product. Historically, these four functions were bundled because retail customers were shielded from real-time prices and network constraints made it difficult for grid operators to selectively curtail customers with a lower willingness to pay for firm capacity.

Yet as we have argued, bundling these four functions together creates a number of challenges for regulators and grid operators. For example, the grid operator estimates the value that LSEs and large loads place on firm service instead of allowing customers to express their individual willingness to pay. Attempting to recreate the operational performance incentives of full-strength energy prices through capacity market non-performance penalties raises significant challenges and potential inefficiencies. Finally, centralized capacity markets make it difficult for market participants to allocate financial risk based on their individual risk tolerances, as PJM assigns capacity obligations to LSEs. One of the principle advantages of Path C is that it reduces many of the regulatory challenges that arise from bundling together these different functions and thus better allocates operational and financial risk.