Evolving Capacity Markets in a Modern Grid

By Robert Stoddard

State policies in New England mandate substantial shifts in the generation resources serving their citizen's electrical needs. Maine, for example, has a statutory requirement to shift to 80 percent renewables by 2030, with a goal of reaching 100 percent by 2050. Massachusetts' Clean Energy Standard sets a minimum percentage of renewables at 16 percent in 2018, increasing 2 percentage points annually to 80 percent in 2050. Facing these sharp departures from business-as-usual, policymakers raise a core question: are today's wholesale market designs able to help the states achieve these goals? Can they even *accommodate* these state policy resources? In particular, are today's capacity markets—which are supposed to guide the long-term investment in electricity generation resources—up to the job?

Capacity markets serve a central role in New England's electricity market. They serve a critical role in helping to ensure that the system operator, ISO New England (ISO-NE), will have enough resources, located strategically on the grid, to meet expected peak loads with a sufficient reserve margin. How capacity markets are designed and operate has evolved relatively little since the New York Independent System Operator (NYISO) launched the first full-fledged capacity market in 1999, the primary innovations being a longer lead-time in procurement of resources, better to support the orderly exit and construction of resources, and higher performance requirements, better to ensure that resources being paid to be available are in fact operating when needed.

Are capacity markets now irrelevant? No. I and many other economists have spent much of the past 20 years working to refine and improve capacity markets with the goal of ensuring that they meet the grid's reliability needs as cost-effectively and efficiently as possible. I believe that they do so—but we must carefully unpack two key terms, "reliability needs" and "cost-effective." The reliability needs of a grid composed principally of conventional generators, able to start and operate on controllable schedules, are very different than reliability needs when intermittent resources like wind turbines and solar arrays account for a large portion of energy generation. Likewise, the definition of "cost" and what "effect" is desired needs to change to reflect state policies. Once we transparently align product definitions and payment streams with state policies, capacity markets will continue to serve in their core function of facilitating efficient investment in a reliable power grid that simultaneously meets states' energy policies.

Capacity Market Basics

Capacity markets are, as Prof. Bill Hogan quipped, the suspenders of a belt-and-suspenders approach to power market design. In theory, expected revenue from sales of energy, ancillary services and (when applicable) environmental attributes should ensure sufficient investment in generation to meet grid needs. Why? Because energy and reserves prices run very high when supplies are scarce, ranging above $1,000/MWh to as high as $10,000/MWh in some markets. If investors see the potential for frequent shortfalls, they should build new resources to capture that scarcity premium. Conversely, in a market with oversupply, generation owners will retire unprofitable resources. This ebb and flow of investment based on market expectations of future scarcity premiums should lead the market to a natural equilibrium of supply and demand—the "belt" of the belt-and-suspenders.

In practice, however, there are three problems with this argument, even in conventional markets dominated by fossil-fueled generation:

- First, various factors may artificially suppress the frequency of realized high prices, such as bid mitigation by market monitors, units dispatched for reliability that are disallowed from setting the clearing price, and *ex post* litigation clawing back supposedly excess earnings.
- Second, this "natural equilibrium" level of capacity reserves may be lower than the installed capacity reserve level that is judged prudent by market operators and regulators. While purely market-driven allocation of investment in other sectors—say, strawberries or hotel rooms—works well, a shortfall of strawberries or hotel rooms has limited impacts on the rest of economy. By contrast, major shortages on the power grid can be highly disruptive and costly.
- Third, investing large sums in a generation facility to chase uncertain scarcity pricing is risky business, and high risk increases the cost of money. One could imagine, in response to the second problem, shifting the "natural equilibrium" reserve margin up by raising the price cap in a market. While this would raise the *expected* earnings (assuming the first issue above is addressed), the economic returns are still concentrated in a just a few hours per year. Spreading payments into more hours reduce generators' risk and, therefore, the required rate of return on investment. This reduction, in turn, lowers consumer costs.[1]

Capacity markets were created to address these three problems while (largely) preserving the incentive effects created by locational marginal prices for energy and reserves. From the 30,000-foot level, capacity markets are fairly straightforward. The market operator determines a target for the quantity of capacity resources for an upcoming year, based on expected peak loads and detailed engineering studies. When there are major transmission constraints in the region, the operator may also specify targets for locational reserve margins. For example, the NYISO targets having enough resources within New York City's peak load to meet 82.7% of the city's expected peak usage. The market operator then conducts an auction to procure capacity commitments from qualified resources to meet these overall and local requirements.[2] The auction clears at a capacity price (or prices, if locational constraints bind) set by the bid of the marginal supplier(s), which in turn should equal that supplier's expected shortfall between the resource's avoidable costs and expected revenues earned in the primary markets (energy, ancillary services, and environmental attributes). Capacity resources have certain obligations, including bidding energy into the market and delivering that energy when called on by the system operator.

[1] A counter-argument is that consumers can contract with suppliers to reduce the risk that they each face in a volatile energy market. In most retail-choice states, however, consumers have the right to change suppliers at least annually and few have the interest or incentive to shop for electricity suppliers.

[2] The three northeast markets all procure capacity with a downward-sloping demand curve, so they may under-procure resources if price is high or over-procure if price is low.

Failure to perform incurs financial penalties, helping to ensure that, by design and maintenance, capacity resources will have high availability when needed most to serve peak demand.

How does this design address the three problems with energy-only markets? In an energy-only market, the installed capacity margin is an *output* of the market design; in an energy-and-capacity market, it is an *input*. When installed reserve margins are above the level that would be supported by energy earnings only, earnings will be lower than they would be in an energy-only market as extra supply drives down prices. Capacity payments top-up these payments to offset this "missing money" between the expected earnings from sales of energy and reserves (and, as applicable, environmental attributes) and the required earnings to keep enough resources available to the market. These payments can be likened to an option payment: capacity resources are induced to maintain and offer extra supply by capacity payments, as well as being paid should the option be exercised. And because these capacity payments are a reliable, steady source of revenue, they reduce investor's exposure to price volatility and, consequently, lower the investment risk premium they must charge.

The Clash Between Capacity Markets and Public Policy

Capacity markets are technology neutral, so it may seem unlikely that some regulators view capacity markets as a block to cost-effective implementation of renewable energy policies. Indeed, when we set up these markets, neither the economists nor the regulators anticipated any such clash. What are the sources of the tension, then, between markets and public policy?

A central issue is a conflict between the form of subsidies to state policy resources, on the one hand, and market power mitigation, on the other.[3] Competitive offers of capacity should be priced to cover the gap between expected costs and revenues, and market monitors are required to assess the competitiveness of offers against this standard. The challenge is *which* costs and *which* revenues to include.

- Allowed costs are those that are *avoidable*. For existing units avoidable costs are generally small, including taxes and operational costs, unless a major refit is needed.[4] For a planned resource, however, avoidable costs also include construction and financing costs. Thus existing resources start from an advantaged cost position relative to new resources—and when state policies are trying to replace fossil-fueled generation with renewable generation, this structural cost advantage in the capacity markets works against that transition.
- Allowed revenues are only those *market* revenues that are available to all similar resources, including energy, ancillary services and Renewable Energy Credits (RECs). This rule is consistent with the Federal Power Act's requirement that wholesale tariffs are not unduly discriminatory. A generator may have a contract on favorable terms, but its capacity offer must be based on the market value of its output, not the contract value. At the time, this rule was adopted to forestall the exercise of buyer-side market power by states seeking to build new gas-

[3] "State policy resources" encompass any resource that receives subsidies from a state, principally renewable or nuclear units.
[4] Fuel costs and other variable charges are not included because these costs are recovered in the energy or ancillary services market revenues.

fired units to lower local capacity prices, a policy actively considered by New Jersey and Maryland. Its effect today is to raise the mitigated offer prices for contracted renewable resources when those contracts are above the market value of energy and RECs those resources can expect to earn. Such above-market contracts, however, are precisely how innovative, high-risk renewables, such as off-shore wind, are being brought into the market.

Taken together, the resulting “competitive” offer price for many state policy resources is much higher than the market clearing price, typically set by incumbent fossil-fired generation. States are still on the hook for their contracted resources, however, leading to concerns that customers are being required to “pay twice” for resources—and, worse yet, are supporting the ongoing operations of fossil-fired units that policymakers would prefer retire.

Today’s markets also carry in their design certain structural biases towards conventional units. These biases exist in both the primary and capacity markets. In ISO-NE’s reserves markets, for example, regulation can only be carried by resources that can generate above or below their generation set point, which effectively excludes wind or solar resources (which can ramp down but not up) and conventional resources operating at their minimum generation (which can ramp up but not down). As another example from the capacity market, ISO-NE’s Capacity Performance (CP) metrics deliver stiff financial penalties against capacity resources that are not generating during super-peak hours (and bonuses for generating above their capacity supply obligations in those hours). These CP penalties give conventional generators a strong incentive to insure units’ availability and fuel supplies. Wind and solar generators, however, are not able to make winds blow or clouds depart. An expensive solution is to install energy storage alongside these intermittent generators, but such grid-scale storage would create a higher social value if it were independently dispatchable by the market operator.

Steps for Improving Markets to Integrate State Policies

The current market design in New England is not without issues, but as a framework it successfully strikes a balance between the “let the chips fall where they may” approach of ERCOT’s “energy-only” market, on the one hand, and the utility-based integrated resource plans that most New England states abandoned nearly two decades ago. The former provides no assurance of resource adequacy, while the latter places substantial risks onto consumers, risks now borne by independent power producers that relate, *e.g.*, to the choice of technology and location. Abandoning the region’s market approach to electricity is premature, unnecessary, and likely to result in the same burdens on consumers that led the region to restructure its electric utilities.

What is needed, I believe, is *more* use of markets, not less, coupled with changes in our markets to allow broader and more complete participation of the evolving technologies that will shape our grid in the next half-century.

1. *Concentrate value in primary markets*

Capacity markets are very good in encouraging investment in resources that have high value in the primary markets, *i.e.* energy, ancillary services and environmental attributes. Regulators and market operators should therefore first focus on ensuring that the attributes they value—for either policy or operational reasons—are seeing this value in primary market prices.

ISO-NE markets are constantly being reviewed and improved, and this process should continue with a focus on monetizing these valuable attributes. For example, PJM is working on a major update to its reserve pricing model, incorporating an administrative demand curve that places greater value on operational reserves during scarcity conditions. Such a change will benefit flexible resources and encourage entry of more providers here—particularly energy storage systems and controllable demand response, which many observers will be critical to expand in order to integrate increasing quantities of intermittent renewables. Such a change would also reduce the number of hours when unpriced operator actions are taken to ensure real-time reliability with the side-effect of suppressing market prices.

In this same line of thinking, ISO-NE needs to expand the range of resources that are allowed to provide operating and capacity reserves. As mentioned above, separating the Regulation product into Regulation Up and Regulation Down would immediately allow more resources to provide one product or the other. Market rules that limit energy storage participation should be carefully reviewed, removing unnecessary restrictions, such as minimum hours of availability at full power.

2. *Create a transparent market price for all clean-energy resources*

As discussed above, a key source of the clash between public policy goals and operation of competitive markets are the growing number of state-backed contracts with renewable resources. Why exactly are policymakers turning to contracts, rather than relying on the operation of the competitive markets?

Contracts serve two important, distinct functions.

- Contracts reduce uncertainty for both buyer and seller; optimally, they shift risk to the party best able to bear it and, in so doing, lower the costs and improve the benefits to both parties.
- Contracts can create and transfer value that is otherwise missing from a market.

For example, a business contracts with an communications provider is principally serving the first, risk-management function, giving the provider a reliable revenue stream to support its investments and the business a stable communications channel. The same business, though, might contract with a specialty manufacturer to produce a custom sub-assembly for its product—an example of a contract that spans the second function, as it would be difficult to create a market for a particular, proprietary product.

Power Purchase Agreements (PPAs) that are primarily of the risk-management sort are typically priced near market values—that is, the *contracted* revenue to the supplier is similar to the *market* value of the energy. The value created by the contract is that the *variance* of the contracted revenues is much lower than the variance of market revenues, which can benefit both buyer and seller.[5] But as we see in the

[5] This statement is particularly true for renewable energy resources, which have no variable fuel costs. Several independent power producers were put into bankruptcy, however, because they had sold power forward at a fixed price but not hedged their natural gas supply. Reducing variance of revenues does not always reduce risk of profits.

contract reviews conducted by the market monitors for the Forward Capacity Market, most of the PPAs with state policy resources are priced well above their expected market revenues, leading to the conclusion that these PPAs are more of the second class of contracts, those providing value otherwise missing from the marketplace.

The major gap is carbon pricing. All six New England states participate in a carbon-pricing mechanism, the Regional Greenhouse Gas Initiative (RGGI), along with New York, New Jersey, Delaware, Pennsylvania and Maryland.[6] The latest clearing price for RGGI allowances, however, was only $5.20, well below any plausible estimates of the social cost of carbon.

RECs could provide an important revenue stream for renewable suppliers, but they have several deficiencies. First, not all state policy resources qualify for RECs—witness the recent Connecticut contracts with Millstone and Seabrook nuclear stations, or the Massachusetts 83(D) contracts with Hydro Québec's large hydro system. Second, RECs do not clearly tie back to the carbon-reducing capability of the generating resource, nor do they provide a focused incentive for energy efficiency, demand response, or distribution-level renewable generation.

The best solution would be adoption of an economy-wide carbon pricing system. There are at least nine such bills in Congress now, but it is beyond the scope of this paper to review their relative merits. Suffice it to say that the best carbon pricing system is the broadest system—broad geographically, to avoid "leakage" across border, and broad in sectors, affecting not just the electric generation sector but also pricing all fossil fuels, regardless of what sector uses them. While political realities may make a national carbon pricing system unlikely in the near term, New England, New York and perhaps other RGGI states form a significant economic bloc that could collectively be an effective carbon-pricing zone, building on the RGGI framework.

Short of this first-best option, New England should adopt a transparent, regional market for zero-emission energy. During the NEPOOL stakeholder process of Integrating Markets and Public Policy (IMAPP), I and other economists presented several variations on the theme of a Forward Clean Energy Market.[7] Central to the operation of such a market is a clear *and common* definition from the states about what, exactly, defines "clean energy." Currently there is a scattershot of legislative definitions for renewable power, differing by vintage, size, and allowed technologies, creating a fragmented and uncertain market for RECs. If the common goal is to reduce carbon emissions, however, it should be possible to craft a common metric for measuring how effectively any technology accomplishes that goal and, therefore, its value in a Forward Clean Energy Market. Such a regional reconciliation and joint action will not be easy, but it is worthwhile.

[6] Pennsylvania and New Jersey had left RGGI and only rejoined in the past few months.
[7] See for example my presentation http://nepool.com/uploads/IMAPP_20161021_CLF_FCM-C_Mechanics.pdf . Other variations and further discussion of this concept are available at http://nepool.com/IMAPP.php.