FCM Performance Incentives
October 2012

Executive Summary

This paper addresses ISO New England's concerns with the performance incentive components of the Forward Capacity Market (FCM). It discusses the ISO's perspective on how the core capacity supply obligations of market participants should change to enhance resource performance and availability.

Empirical analyses of generating unit performance indicate that, at times of high system stress, a significant share of the region's generating fleet fails to respond to ISO dispatch instructions according to their offered capabilities. Reliability risks associated with the region's growing dependence on natural gas-fired generation are compounding these concerns. Many of these challenges could be resolved if suppliers undertook additional operational-related investments, whether in dual-fuel capabilities, short-notice and/or non-interruptible gas supply agreements, new fast-responding demand response assets, or other arrangements to similar effect. However, the present FCM design provides little incentive for suppliers to undertake these investments.

The ISO proposes to modify the FCM design to make each resource's FCM revenue contingent, in part, upon its actual performance during periods when aggregate performance does not enable the ISO to satisfy system reserve requirements. The new performance incentive design will result in transfers from under-performing to over-performing resources, providing strong incentives for each resource to perform as needed and for resources that can meet the system's needs by exceeding their obligation to benefit by doing so. These incentives will place performance risk on all FCM resources, and this risk will need to be priced in each resource's bid in future capacity auctions.

The proposal structures the transfers among suppliers so that consumers continue to pay the forward capacity auction clearing price; consumers will not bear the short-run risk of covering unexpectedly high performance incentives. This will continue to provide consumers with a predictable capacity price three years out, after the close of each forward capacity auction. In addition, the performance incentives should reduce the need to specify explicit flexibility requirements in the FCM.

The first half of this paper summarizes the economic framework underlying the FCM, explains why energy markets provide insufficient performance incentives during scarcity conditions on the power system, and identifies the essential features of economically sound FCM performance incentives. The second half of the paper lays out the proposed pay-for-performance approach, with sufficient detail to facilitate productive discussions about the approach with stakeholders.

These changes will require a significant amount of time and effort from the region, and the ISO looks forward to reviewing this paper with stakeholders as a first step in that process.

1. Introduction and Overview

In 2010, ISO New England launched a Strategic Planning Initiative to focus the region on developing solutions to five challenges confronting New England's power system and wholesale markets.[1] The first challenge concerns resource performance and flexibility. This challenge arises from a growing concern that, as the power system continues to evolve, the emerging mix of supply resources may be unable to operate when and as needed to maintain the grid's present level of reliability. Significant changes to performance incentives in the FCM are needed to resolve these concerns.

In New England, concerns with resource performance and flexibility arise from several sources. One risk is the operational performance of existing resources during stressed system conditions—times when resources' performance is essential to reliability. ISO analyses indicate that older units that are relied upon for peaking service, ramping, or reserves are not performing within their offered parameters.[2] These shortcomings became manifest in operational events on June 24, 2010, September 2, 2010, and January 24, 2011 (including a NERC violation related to inadequate generation contingency response on September 2).[3] More generally, an examination of dispatch response performance following the 36 largest system contingency events over the last three years indicates that, on average, New England's non-hydro generating fleet delivered less than 60% of the additional power requested of these resources by the ISO. In sum, at times of greatest need, many resources are delivering far below the performance ability represented in their supply offers.

A second source of growing concern stems from New England's increasing reliance on natural gas-fired generation.[4] Gas-fired resources rely upon a "just in time" fuel delivery system using inter-state pipelines that, in general, must be scheduled in advance of the operating day to ensure adequate fuel. Whenever New England's power system experiences an unforeseen problem in the natural gas supply chain during the operating day, it must be able to rely on flexible resources that have fuel to maintain system reliability. As the region's reliance on natural gas expands, greater private investment in hardware, fuel arrangements, or other supplier-selected solutions to ensure resource availability is essential. Changes to the FCM can improve participants' incentives to undertake these investments.

A third concern with resource performance and flexibility arises from changes in the mix of supply resources that participate in New England's energy markets. Specifically, the potential growth in intermittent power sources will create a greater need for flexible supply resources to smooth intermittent resources' fluctuations in output during the operating day. The ISO identified this need within the Strategic Planning Initiative effort to integrate higher levels of intermittent resources into

---

[1] For an overview, see *Roadmap for New England: A Proposal for Meeting the Challenges Identified in the Strategic Planning Initiative* (March 2012), at http://www.iso-ne.com/committees/comm_wkgrps/strategic_planning_discussion/materials/strategic_plan_initiative_roadmap_march_2012.pdf.

[2] *Strategic Planning Risk Summary* (21 April 2011), p. 4. At http://www.iso-ne.com/committees/comm_wkgrps/strategic_planning_discussion/materials/spd_risk_summary_apr_2011.pdf.

[3] *Ibid,* p. 4.

[4] See *Addressing Gas Dependence* (July 2012), at http://www.iso-ne.com/committees/comm_wkgrps/strategic_planning_discussion/materials/natural-gas-white-paper-draft-july-2012.pdf.

the grid.[5] Changes to the FCM that improve incentives for resource flexibility and availability will provide better incentives for investment in resources that can balance intermittent power supply. While this strategic planning risk is a longer-term concern, addressing FCM changes now to improve incentives for resource flexibility will facilitate investment planning decisions by the private sector.

Incentives

Connecting many of the challenges identified in the Strategic Planning Initiative is a common concern: The physical performance incentives in the region's wholesale electricity markets are inadequate. This is particularly true for the FCM. In principle, the FCM can and should provide economically sound, market-based incentives for a capacity resource to supply energy when needed to maintain system reliability. Addressing performance incentives via the FCM is paramount in a market environment where wholesale energy prices, even during operating reserve deficiencies, reflect short-run marginal costs rather than the value New England places upon reliable electric service. Without greater performance incentives, resources' availability—and investment in technology, fuel arrangements, and business practices that maximize their availability—will be insufficient to maintain the grid's reliability in the future.

Consumers may prefer a certain level of insulation against difficult-to-predict changes in FCM costs. Providing this insulation will require a careful balance in the design of performance incentives. By necessity, incentives entail risk; without the 'upside risk' of reward for strong performance, and the 'downside risk' of lower revenue for poor performance, desirable performance and investment incentives cannot be achieved and the reliability risks to New England's power system will remain. As discussed below, the proposed approach to performance incentives balances the twin objectives of an economically efficient risk-and-reward structure for FCM suppliers, and relatively predictable total capacity cost for New England consumers three years hence.

Approach

This paper describes how the core capacity supply obligations of market participants must change to improve resource performance incentives. The next section provides a conceptual discussion of the economic incentives that the FCM should provide to achieve essential changes in resource performance and investment. This discussion identifies a number of market design principles that are inconsistent with the existing FCM rules. These inconsistencies lead to the conclusion that simply increasing the severity of the existing FCM penalty structure will not achieve an efficient, reliable system.

Rather, the existing FCM Shortage Event penalty structure must be replaced by a new pay-for-performance mechanism. A central component of this mechanism is the creation of strong financial incentives for all capacity suppliers, without exception, to perform during scarcity conditions. Scarcity conditions, in this context, occur any time the ISO is unable to meet the combined energy and operating reserve requirements necessary to ensure reliable operations. A second component of the pay-for-performance mechanism is that the FCM's economic incentives should include financial

[5] *Strategic Planning Risk Summary* (21 April 2011), p. 7. At http://www.iso-ne.com/committees/comm_wkgrps/strategic_planning_discussion/materials/spd_risk_summary_apr_2011.pdf.

transfers from under-performing resources to over-performing resources during these scarcity conditions. This means consumers will not bear the short-run risk of covering unexpectedly high performance incentives.

Last, a consequence of stronger incentives is that capacity suppliers will face greater financial risk. As we explain in this paper, this will motivate suppliers to take actions to reduce financial risk by improving resources' physical performance. Many suppliers may incur new costs to provide more reliable service. This will impact future capacity auction prices, as suppliers will reflect these costs and financial performance risks in their capacity auction bids.

Benefits

The FCM performance incentives described in this paper will provide four important benefits.

► **Operational-Related Investment.** Strong performance incentives provide suppliers with the economic motivation, and the financial capability, for operational-related investments that ensure resources are available when needed to maintain reliability. Some of the actions that may be incented are dual-fuel capability, short-notice or more reliable fuel supply arrangements, continuous staffing at resources, rapid price-responsive demand behavior, and other improvements to similar effect.

► **Increased Resource Responsiveness and Flexibility.** Improved performance incentives should lead suppliers to revise their operating procedures to maximize availability and responsiveness to ISO instructions, which are essential to ensure reliability. This may be achieved through improved operating practices, incremental capital investments that shorten start times or increase ramp rates, and a change in the resource mix over time as high cost, inflexible resources are replaced by low cost and more flexible resources.

► **Cost-effective Solutions.** Markets motivate suppliers to deliver services in the most cost-effective ways. Performance incentives will enable individual suppliers to select the solutions that work best for the technologies and features of their resources. This market-based approach rewards suppliers that pursue the most cost-effective means to improve performance and availability.

► **Efficient Resource Evolution.** Finally, stronger performance incentives will, over time, lead to a change in the capacity resource mix that directly improves system reliability at lowest cost. Resources that are unreliable and have high operating costs may submit higher offers into the Forward Capacity Auction (FCA), based on their expectation of performing poorly and experiencing non-performance penalties during the commitment period. These resources will become less likely to clear the auction, relative to today. In contrast, the compensation provided for strong performance will enable highly efficient or highly flexible resources to profitably make lower offers in the FCA, and they will therefore be more likely to clear future capacity auctions.

2. Background: FCM Design

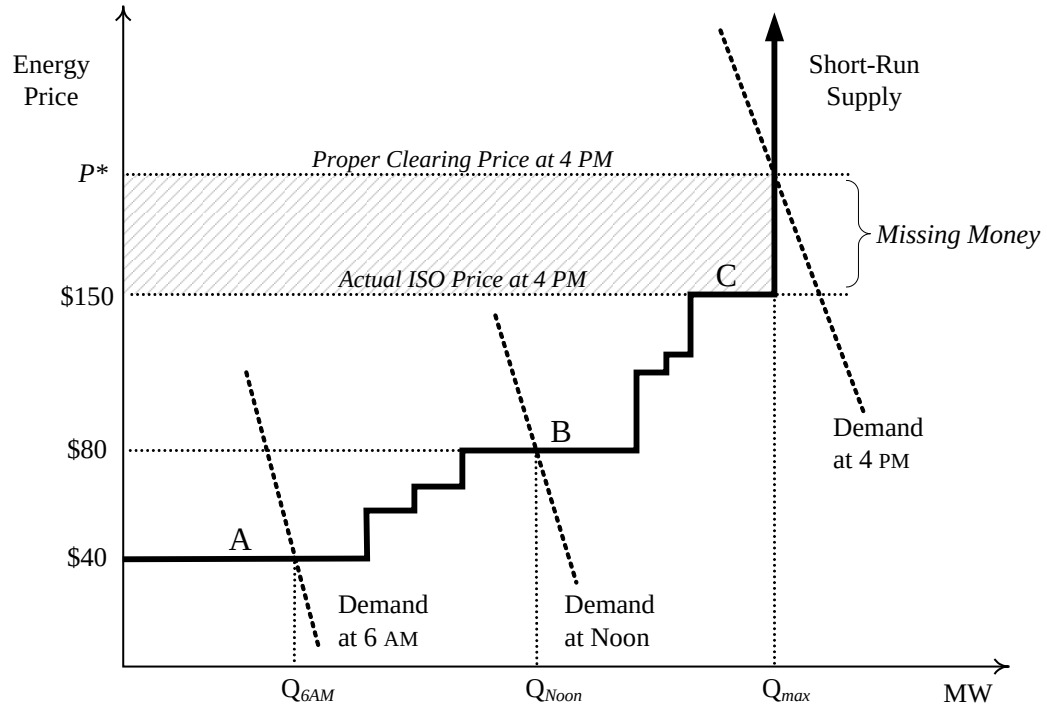
The logic of FCM performance incentives is linked closely to the central purpose of a capacity market: To solve the "missing money" problem that occurs in energy-only electricity markets. In this section, we summarize the economic framework underlying the FCM. The purpose of this discussion is to identify the essential features of economically sound performance incentives. These features lend themselves to a pay-for-performance design that differs from the existing FCM penalty structure.

Concepts

The economic logic of the "missing money" problem merits a brief explanation. Figure 1 depicts a simplified market-level supply curve in a competitive, short-run electricity market. Each 'step' represents the marginal cost and quantity of a different supply resource. For the sake of clarity, we ignore transmission constraints, operating reserves, and other engineering complications.

The performance and investment problems with energy-only electricity markets arise from the way prices are set when demand reaches the market's short-run capacity limit. Consider demand levels that change over the course of the day, as illustrated in Figure 1. At 6 AM, demand is low and supplier A sets the market-clearing energy price at $40 per MWh. At noon, demand is higher and supplier B sets the market-clearing energy price at $80 per MWh. Note that at the noon hour, supplier B is the marginal producer and therefore earns zero net revenue in the energy market.

FIGURE 1. A Short-Run Energy Market



What happens when demand is highest, at 4 PM? First, consider what the efficient price *would* be in an idealized energy market. When demand reaches the market's short-run capacity, the market-clearing price would be where supply intersects demand, as always. In Figure 1, this is price level p^*. Note that when the market's short-run capacity limit binds, the efficient price p^* is *not* equal to any supplier's short-run marginal cost. If the price was set at marginal cost (here, $150 per MWh), demand would exceed capacity (the quantity Q_{max}) and there would be a shortage—some demand would go unserved.

Outside of electricity markets, this 'idealized' market clearing works smoothly in many different settings. For example, consider industries that have both short-run capacity constraints and provide a service, or produce a limitedly storable good: Hotels, oil refining, airline flights. In those markets, price rises to a level analogous to p^* in Figure 1 when the industry's short-run capacity limit is reached. When demand is lower and sellers find they have idle capacity, unbooked hotel rooms, or unsold airline tickets, price falls closer to marginal cost.

What purpose is served by having price rise to p^*? In many markets, the periods when short-run capacity limits bind and price rises above marginal cost provide sellers with the only opportunity to cover their total costs (including a return on capital invested). In the context of the idealized energy market illustrated in Figure 1, marginal supplier C must cover all of its fixed cost (including its cost of capital) from the revenue it receives in the few hours when demand reaches total capacity, Q_{max}.[6] Put another way, the revenue from high spot electricity prices at times when demand reaches short-run capacity serves an essential purpose: Without it, marginal suppliers could not expect to recover their total costs, and would not enter the marketplace (or will soon exit). In that event, additional demand would go unserved. If these high but efficient prices are not present at the appropriate times to spur investment, then a different market mechanism must be developed to ensure investment.

Energy Market Prices and "Missing Money"

It is useful to connect the logic of this idealized market to the realities of wholesale power market prices. In practice, the ISO does not observe price-sensitive (that is, downward sloping) demand in the real-time electricity market. Without that information, the electricity market cannot set price at p^* when capacity limits bind, as it theoretically should and as occurs in other capital-intensive industries. Instead, electricity market operators set price at the cost of the marginal resource even when demand reaches the market's short-run capacity limit.[7] In context of Figure 1, this means suppliers may be paid a price equal to the incremental energy cost of marginal supplier C, or $150 per MWh, at times when the price that should prevail, in theory, is the efficient market-clearing price of p^*.

---

[6] It is occasionally suggested that in capital-intensive industries, producers must be able to exercise market power in order to recover their fixed costs. That is not correct. As shown in Figure 1, a proper market-clearing price of p^*, which is above short-run marginal cost, is not consistent with market power: No supplier is withholding output in order to raise price.

[7] In practice, electricity market operators can set price above marginal cost when operating reserves are deficient, at an administratively determined shortage price (known as a "Reserve Constraint Penalty Factor," or RCPF). This practice can reduce the magnitude of the missing money, but in New England the RCPF was not developed to approximate the efficient market-clearing price p^*.

Market participants occasionally point to regulatory offer caps on suppliers as the cause of this pricing problem. Offer caps contribute to the problem: If price is administratively mitigated to the incremental cost of the marginal supplier, then there will be missing money in the energy market at times when demand reaches short-run capacity. Offer caps are not the only cause of the missing money problem, however. Even without offer caps, the absence of sufficient price-sensitive demand behavior in the spot (real-time) electricity market means the ISO cannot determine the proper market-clearing price, p^*, as depicted Figure 1. Instead, there is only a single value of (real-time) demand, which is insufficient to determine the value of p^*.

To date, this problem has proven difficult to solve in energy-only electricity markets.[8] It results in under-pricing of electricity at times when demand reaches short-run capacity, and yields inefficient incentives. The inefficiency has material consequences: It undermines investment in generation and demand resources that are necessary to maintain a reliable power system. In effect, suppliers are "missing" a revenue stream that is necessary for them to cover their total costs. This revenue stream is known as the "missing money" that energy-only electricity markets do not provide, but must replace—by some means—in order to prevent suboptimal investment in the power system. [9]

FCM Objectives

The central objective of the FCM is to create a revenue stream that replaces the missing money, and thereby provide appropriate financial incentives to invest in capacity. Setting aside (for the moment) some of the complex details of the FCM, the basic logic of how a capacity market replaces the missing money is straightforward. A forward capacity auction determines the minimum contribution to each potential supplier's total costs that the supplier would be willing to accept to enter, or to not exit (as applicable). The auction clears the least-cost set of bids that satisfy (forecast) demand. The resulting capacity auction clearing price replaces, in expectation, the missing money revenue stream.

Although capacity auctions can identify the missing money amount, a crucial second problem must be addressed. The under-pricing of electricity at times when demand reaches short-run capacity not only creates missing money, it also undermines suppliers' incentives for resource performance and availability. This occurs because at times when capacity constraints bind—and the need for a supplier's energy is greatest—suppliers are paid marginal cost (or an administratively set shortage price), which is less than the proper market clearing price.

This is an incentive problem that competitive markets normally solve, but energy-only electricity markets do not. The solution requires aligning a supplier's economic incentive for production with

---

[8] In principle, this problem could be addressed by setting a sufficiently high administrative shortage price in the energy market. That approach seeks to approximate the efficient price p^* during scarcity conditions, and could provide suppliers with similar incentives to the FCM performance incentives described below. However, a high administrative shortage price in the energy markets would increase energy price risk for both consumers and producers, unless it is coupled with a carefully-designed risk sharing mechanism (analogous to an expanded Peak Energy Rent deduction in New England's FCM).

[9] A different form of "missing money" can also occur if administratively determined installed capacity requirements exceed the level of capacity that would be selected by an efficient market. If so, an energy-only electricity market's revenue may not induce a level of investment that meets the administrative capacity requirement. The forward capacity auction also solves this form of the missing money problem.

the value consumers place upon reliable service at times when demand reaches short-run capacity. This is the FCM's second objective, and it presents additional market design challenges beyond the conduct of capacity auctions.

Market Design Principles

Done properly, the way the FCM pays out the missing money can replace the strong performance and availability incentives that properly functioning markets normally provide. Three aspects of these strong performance incentives have particular importance, as they are essential principles of efficient markets.

► **Markets Pay for Performance.** The first observation is that markets naturally possess an effective pay-for-performance mechanism. Specifically, a properly functioning market will pay a supplier (the equivalent of) the missing money revenue *only to the extent it is producing* at the time.

In the context of Figure 1, for example, consider the missing money revenue represented with the shaded rectangle. In an efficient market, suppliers can earn this revenue because they receive the market-clearing price p^* during periods when demand reaches market's short-run capacity. Accordingly, to mimic an efficient market, the missing money paid through the FCM should be similarly contingent upon a supplier's production at times when demand reaches short-run capacity. The pay-for-performance approach described further below satisfies this key market principle.

► **Fault and Financial Risk**. The second feature of efficient incentives is that they place risk upon a supplier. Specifically, a supplier may not be able to cover its total costs if it is unavailable when demand reaches the market's short-run capacity. Placing that level of risk on suppliers is precisely how an efficient market works, and why it creates strong financial incentives for resource performance. These strong incentives are important because they motivate suppliers to produce, to their utmost capacity, at times when consumers value it the most and the system requires it to run reliably.

In effect, markets are blind to fault. If a supplier is not producing—for any reason—at the time demand reaches short-run capacity, it misses out on the opportunity to receive the market's (equivalent of the) missing money. It does not matter why the supplier is not producing, or whether the reason(s) are within, or beyond, the supplier's control. Those are business risks that suppliers must manage, and their entry and exit decisions—and expected market prices—should reflect that operational risk.

An efficient FCM pay-for-performance approach should mirror these fault and financial risk principles. Accordingly, if a resource does not produce at times when demand reaches short-run capacity, it should receive less of the missing money revenue—irrespective of why it does not produce. This places greater financial risk on suppliers than today. Each supplier will perceive this risk differently, and should be expected to reflect this risk in its future capacity auction bids.

► **Resource Neutrality.** The third important principle of competitive markets is that two suppliers that provide the same good or service receive the same price.[10] Their compensation is not dependent upon whether they use the same means of production.

In effect, an efficient market provides the same incentives to all suppliers, regardless of resource type or technology. Accordingly, efficient FCM performance incentives should provide obligations to perform, and payment terms for performance, that are the same for all suppliers irrespective of technology. This will ensure comparable treatment for comparable performance.

Implications

Taken together, these observations highlight the way that competitive markets provide incentives to perform at times of greatest need: Suppliers earn the missing money revenue stream by producing when demand reaches short-run capacity. As noted earlier, however, an energy market that sets price equal to marginal cost cannot replicate this efficient pay-for-performance system. The logical conclusion is that, in order for New England's wholesale market system to achieve the benefits of competitive markets and their strong incentives for performance, the FCM's market design must change to replicate these economic incentives.

Replicating the performance incentives of properly functioning competitive markets does not require fundamental changes to the auction elements of the FCM. The initial capacity auction (and reconfiguration auctions) serves as the device to determine the missing money—the potential reward—earned by suppliers that perform. However, the terms under which suppliers earn the missing money must be contingent upon performance in ways that mirror an efficient market. To do so will require re-defining the conditions under which FCM suppliers are paid the missing money revenue stream.

3. FCM Performance Incentives: When and Why

While the foregoing discussion describes how properly functioning markets provide strong performance incentives, it entails some simplifications that must be clarified. These include the role of operating reserves and what situations constitute "scarcity" conditions. This section clarifies the conditions under which FCM performance incentives should apply. It also explains why improving FCM performance incentives during these conditions will change resources' operational and investment decisions, ensuring reliability in cost-effective ways.

Scarcity Conditions

The economic insights illustrated in Figure 1 are based on a simple notion of demand reaching a single, short run "capacity" limit. Power systems are more complex. Economic theory suggests, as illustrated in Figure 1, that the FCM should provide additional performance incentives during

[10] This property is known as the law of one price: "In any market, at any moment, there cannot be two prices for the same kind of article." Jevons (1871), *The Theory of Political Economy*.

conditions—and only during conditions—when the energy market price is below the efficient market-clearing price. These situations are known as *scarcity conditions.*[11]

In electricity markets, scarcity conditions occur at times when the power system is deficient *any* of its operating reserves. This is a specific, well-defined operating condition. The reason it matches precisely when scarcity conditions occur is important, and involves three observations about the power system.

First, an operating reserve deficiency occurs when total power supply available to the system in real-time is insufficient to satisfy the sum of load (*i.e.*, energy demand) and the system's operating reserve requirements.[12] Operating reserves provide essential protection against unforeseen contingencies (such as sudden generator or transmission equipment failures); without these reserves, the system could not operate without disconnecting customers involuntarily.

Second, consider what the efficient market-clearing energy price *should* be when the system is deficient operating reserves. In an ideal energy-only electricity market, the market-clearing price would rise above marginal cost, to a level where demand would fall by (just) enough that some available supply resources can be used to eliminate the reserve deficiency. This can be achieved because if demand falls, the ISO is able to reallocate resources in real time from supplying energy to providing reserves. By doing so, the system would maintain the physical insurance necessary to protect against contingencies that otherwise would threaten overall system reliability.

Third, the efficient market-clearing price differs from the energy price that prevails in practice. During scarcity conditions, the ISO cannot balance demand with the limited supply of energy while maintaining required reserves. As discussed previously, the energy market continues to set price at either the incremental cost of the marginal energy supplier or an administratively set shortage price (RCPF). This results in a price gap—between the efficient price (p^*, in Figure 1) and the actual price paid to suppliers—whenever the power system is deficient operating reserves.

The central point to observe is that scarcity conditions arise whenever the power system is deficient any of its operating reserves. These are the times when, in a properly functioning market, suppliers would have the opportunity to earn (the equivalent of) the missing money. Accordingly, to replace the strong performance incentives that are missing from the energy market, the FCM must provide suppliers with comparable performance incentives during—and only during—real-time conditions when the power system cannot meet its operating reserve requirements.

---

[11] The difference between the efficient market-clearing price and the marginal supplier's incremental cost, when positive, is also known as a *scarcity rent.* The relationship between scarcity rent and scarcity conditions, as used in this paper, is precise. Scarcity conditions occur when the price in the actual energy market does not yield the scarcity rent that a supplier should earn in an ideal, energy-only electricity market.

[12] As a technical note, the ISO may experience an operating reserve deficiency if there are no additional generation resources physically available to provide reserves, or if the incremental cost of additional operating reserves exceeds certain limits (the reserve constraint penalty factors). At present, these limits are set such that both conditions generally occur simultaneously.

Prevalence

In recent years, the New England power system has experienced system-level operating reserve deficiencies for a total of 29 hours per year, on average.[13] Recent and pending changes to ISO operating procedures are likely to change this in the future. Specifically, certain actions that the ISO initiated *after* entering a reserve deficiency in the past will be initiated *prior to* a reserve deficiency in the future. For example, with the full integration of real-time demand resources into the energy markets in 2017, these resources will be dispatched economically before entering a reserve deficiency; today, these resources are dispatched as a curative action after a reserve deficiency commences. In addition, the ISO increased the system RCPF for total (thirty-minute) operating reserve from $100 to $500 per MWh in June 2012. The change was implemented to improve both reliability and the accuracy of energy and reserve price signals in the real-time markets. This change will reduce the frequency of operating reserve deficiencies by dispatching additional resources, in economic merit order, prior to entering a reserve deficiency.

These changes have consequences for the prevalence and significance of scarcity conditions. If New England continues to have excess capacity, the total annual hours of scarcity conditions may be slightly lower going forward than in recent years' data. While that may appear helpful from a reliability standpoint, there is a caveat. In the future, when an operating reserve deficiency occurs, system operators will have fewer *remaining* options at their disposal to cure it. This means that when the power system enters a reserve deficiency in the future, it may be in a more severe condition than typically was the case in the past. As a consequence, it is essential to have strong performance incentives that minimize the severity of future deficiencies and ensure the system's recovery to normal operating conditions.

Why Improving Incentives Changes Performance

One of the virtues of markets is that efficient prices motivate performance by suppliers in the most cost-effective ways. Similarly, providing FCM suppliers with strong performance incentives during scarcity conditions will not only improve system reliability, it will motivate suppliers to do so at minimum cost using different means.

Strong performance incentives will lead suppliers to revise their business practices to maximize their ability to supply energy and reserves during scarcity conditions. Operational practices that enhance this ability include reducing a unit's startup and notification times, improving its dispatch response following system contingencies, keeping additional staff at power facilities or demand-response control centers to ensure the resources are available when system conditions are tight, and so on. Of course, some suppliers already do all of these activities. Still, nearly any operational practice that improves a resource's availability and response flexibility entails some cost, and the extent to which suppliers undertake them depends upon their business analysis of whether the market's anticipated compensation exceeds the cost. Creating new performance incentives during scarcity conditions will change these benefit-cost calculations, rewarding improvements in resource availability and performance.

---

[13] Zonal reserve deficiencies without a system-level reserve deficiency are uncommon (occurring less than one hour per year in the last four years).

Beyond operational practices, improved resource availability will also result from operational-related investments. Such investments may be useful only on occasion, but are still cost-effective ways to improve system reliability. For instance, certain operational-related investments would directly address reliability concerns related to New England's growing dependence on natural gas-fired generation and its "just-in-time" fuel delivery system, such as maintaining dual-fuel capability, developing new fuel-supply arrangements with pipeline service providers, and the like.

On these points, an example may help. Imagine the basic business case for the owner of a gas-fired combined cycle facility to install dual-fuel capability that would be used only a small number of hours per year:

- Assume the owner of a combined-cycle generating station expects there to be ten hours per year in which the power system experiences an operating reserve deficiency due to gas pipeline limitations. Suppose that if the station owner installs dual-fuel capability, it will be able to run at full capacity using oil during these reserve deficiency hours; if the owner does not install dual-fuel capability, it will not be able to operate during these ten hours.

- Assume the combined-cycle facility can install dual-fuel capability at a cost of $15,000 per MW-year, including both annualized capital costs and recurring costs of maintenance, testing, and inventory turnover of unused oil. If FCM performance incentives provide similar incentives to those of a properly functioning competitive market, it should pay a high price (analogous to p^* in Fig. 1) for resources that operate during scarcity conditions. If, as a hypothetical matter, the value of p^* is (say) $2,000 per MWh, then it becomes profitable for the combined-cycle owner to invest and to maintain the dual-fuel capability in order to run those ten hours: 10 hours × $2,000 per MWh = $20,000 per MW-year of expected revenue, which exceeds the $15,000 per MW-year in annualized costs.

By contrast, in the absence of strong performance incentives during scarcity conditions, the combined-cycle facility owner would not find it cost-effective to install the dual-fuel capability—and the severity of the operating reserve deficiencies would be worse.

Of course, different suppliers may find other operational-related investments a more attractive means of achieving the same result. For instance:

- Assume the same facts, and that the addition of a new compressor on the gas pipeline would enable the pipeline company to offer an improved fuel delivery agreement to the combined-cycle station. The new fuel delivery agreement would provide the station with gas during conditions when, without the new compressor, gas would not be available to the station. In this situation, the combined-cycle owner would choose whichever product—dual-fuel capability or the pipeline's terms for the new compressor—is more cost effective and, in the generator's view, more likely to assure that it will be available for those ten hours.

- Finally, imagine that a new demand response resource can provide the same service, by reducing load on the power system during the ten hours of scarcity conditions. Assume this demand resource can be developed at a lower total cost (per MW) than the cost of adding dual-fuel capability to the combined-cycle station (per MW). In this case, the new demand response resource would be a cost-effective investment even if it operated only during the ten hours of scarcity conditions annually.

In sum, the benefits of markets arise from all the ways in which suppliers will develop cost-effective means to increase their availability and flexibility. This will improve their ability to operate at times when the need is greatest. The span of these cost-effective investments might include innovative fuel arrangements for intra-regional gas storage with local distribution companies, short-notice service agreements developed with gas infrastructure providers, backup fuel supplies, new price-responsive demand arrangements, and so on. Different solutions will appeal to different FCM resources, given their existing infrastructure, location, and operating characteristics.

The point is that the most efficient solutions to the region's reliability requirements will surely come from the innovative results of supplier-selected solutions. These solutions must be motivated by the sound economic incentives that exist in competitive markets, and that must be established in New England's electricity market design.

4. A Pay-for-Performance Approach

The preceding sections establish why the energy market alone provides insufficient performance incentive during scarcity conditions. To resolve this incentive problem, changes to the FCM are required. In this section, we describe a pay-for-performance approach that provides capacity suppliers strong, economically sound, market-based incentives to perform at times of need. The ISO proposes to replace the existing FCM Shortage Event penalty structure, in its entirety, with a new pay-for-performance approach.

The ISO's proposed pay-for-performance approach adheres to several market design principles that characterize efficient, competitive markets:

- It enables suppliers to earn the missing money revenue stream that an efficient energy market would provide, by delivering energy and reserves during scarcity conditions;[14]

- It provides performance payments and charges contingent upon actual performance, irrespective of fault;

- It provides the same incentives to all suppliers, regardless of resource type. Consistent with a competitive market, it neither favors nor discriminates against any class of resources.

Under the ISO's proposed pay-for-performance approach, a resource's total FCM revenue is tied directly to its performance during scarcity conditions. There are three central components to this pay-for-performance approach:

1. Creating strong economic incentives for all capacity suppliers, without exception, to perform during scarcity conditions;
2. Implementing these incentives through transfers from resources that under-perform to resources that over-perform during scarcity conditions; and

[14] The efficiency is approximate, inasmuch as the ISO is obligated to determine installed capacity requirements based on administrative reliability standards. These administrative standards are not expressly tied to consumers' willingness to pay for reliable service or the costs of outages.

3. Defining scarcity conditions as any time in which the ISO is unable to satisfy the combined energy demand and operating reserve requirements of the power system.

The first component involves changing the FCM market design so that resources with superior performance during scarcity conditions are able to earn a greater share of the missing money revenue stream determined by the FCA. Resources with inferior performance during scarcity conditions earn a smaller share of the missing money revenue stream. This is consistent with the central tenets of how an efficient, properly functioning market remunerates (the equivalent of) the missing money to suppliers, as described in Section 2. In effect, it ensures that suppliers face risk and reward for their performance at the right times—the times when the energy market alone provides performance incentives that are too low and, simultaneously, system reliability needs are high.

The second component serves to balance the twin objectives of a sound risk and reward structure for FCM suppliers and a relatively predictable total capacity cost for New England consumers three years hence. It ensures that consumers do not bear the financial risk of unexpectedly high incentive payments earned by high-performing suppliers during the FCM delivery year. Instead, it is the low-performing resources that bear the financial risk. This allocation of risk provides sound economic incentives for suppliers not to under-perform during scarcity conditions.

The third component reflects the central observation that the periods when the energy market provides insufficient performance incentives to suppliers are precisely the times when the power system is deficient operating reserves. This condition is a direct consequence of the analysis in Section 3.

Mechanics of Pay-For-Performance

These objectives and principles of economically sound incentives can be achieved with a conceptually straightforward incentive mechanism. The central idea is that a supplier's FCM revenue comprises two parts: A *base payment*, and a *performance payment*.[15] The base payment is determined by the forward capacity auction result. The performance payment is determined by a resource's performance whenever scarcity conditions occur during the capacity commitment period. A resource's performance payment may be a positive or negative adjustment to its base payment, reflecting superior or inferior performance during scarcity conditions.

Performance Scoring

The performance payment is determined by a resource's *performance score.* The performance score is calculated, separately for each resource, during each interval in which scarcity conditions occur. The performance score is the difference between the resource's actual performance and a share of its capacity supply obligation (CSO):

$$Score = Actual\ \text{MW} - \text{CSO MW} \times Balancing\ Ratio.$$

[15] Note that there are additional elements to a supplier's FCM obligations, such as the Peak Energy Rent deduction and financial assurance obligations. The base and performance components described here do not comprise all of the FCM's financial provisions.

The performance score is measured in MW. It is essential to note that the resource's actual MW are determined by the sum of its energy production and the reserves that it provides at the time. Thus, for purposes of determining the performance score, a resource supplying 100 MW of energy and an additional 50 MW of reserves would have an actual MW value of 150 MW.

The CSO MW is the resource's capacity supply obligation at the time. The resource's CSO MW is adjusted by a *balancing ratio* to account for the total energy and reserve requirement at the time. The balancing ratio is a proportionate adjustment to CSO MW:

$$\textit{Balancing Ratio} = (\textit{Load} + \textit{Reserve Requirement}) \ / \ \textit{Total CSO MW}.$$

For instance, suppose a scarcity condition occurs during an off-peak period when load is 16 GW and the reserve requirement is 2 GW. Assume total CSO MW in the FCM is 30 GW. Then the balancing ratio would be (16 + 2) / 30 = 60%. Thus, each resource's actual performance would be compared to a reference level of 60 percent of its CSO MW during this interval. As an example, suppose a resource has an actual MW value of 150 MW and a CSO MW of 200. If the balancing ratio is 60%, its score is +30 MW for this interval. Alternatively, if the resource's actual MW is 100 MW during this interval, then its score is –20 MW.

The balancing ratio ensures that performance incentives do not penalize resources for failing to deliver their full CSO MW if scarcity conditions occur when the system's total energy and reserves requirements are substantially less than the installed capacity requirement (ICR). Using the preceding figures, if the system's total energy and reserve requirements are only 18 GW at the time scarcity conditions occur, it is not necessary for capacity suppliers to deliver the total CSO MW of 30 GW. However, if energy demand plus required reserves reaches the total CSO MW, the balancing ratio is 100%. In this event, every capacity resource is needed to supply its full CSO MW in the form of energy and reserves.

Importantly, if suppliers provided (in aggregate) more than the 18 GW needed in this example to meet energy demand and reserve requirements, then the system would no longer be in a scarcity condition—and the FCM performance incentives would not be in effect. This prevents suppliers from having an incentive to 'oversupply' as the system emerges from a scarcity condition.

FCM Performance Payments

As indicated above, a resource's FCM revenue would be determined, in part, by two components: a base payment and a performance payment. On a monthly basis,

$$\textit{FCM Payment} \ = \ \textit{Base Payment} \ + \ \textit{Performance Payment}.$$

The resource's base payment component is determined by the FCA clearing price like today:

$$\textit{Base Payment} \ = \ \text{FCA } \textit{Price} \ \times \ \text{CSO MW}.$$

The resource's performance payment is determined by its performance scores:

$$\textit{Performance Payment} \ = \ \textit{Performance Payment Rate} \times \ \textit{Total Score}.$$

There is a performance score for each interval in which scarcity conditions occur. No score is calculated for periods when there is no operating reserve deficiency. Each month, a resource's total score is the duration-weighted sum of its performance scores during the payment period.[16] Thus, if there are two hours of scarcity conditions in a payment period and a resource's performance scores are +30 MW and –20 MW in each hour, respectively, its total score is +10 MW.

The performance payment rate (defined in dollars per MWh), in combination with the real-time prices of energy and reserves, determines suppliers' economic incentives to perform during scarcity conditions. It must be set high enough to ensure that it materially impacts the amount of missing money a supplier earns, as would occur in a normal, efficient market. There are several important economic principles that guide the determination of the performance payment rate, discussed further below.

Examples

Some simple examples help illustrate this pay-for-performance approach.

► **Example 1.** No scarcity conditions in a month.

If there are no scarcity conditions, no performance scores are calculated (all scores are zero). FCM payments are not adjusted for performance. The monthly payment is the FCA Price × CSO MW.

► **Example 2.** One hour of scarcity conditions occur in a month, when load and reserve requirements equal total CSO MW, and a resource with a 100 MW CSO performs at 90 MW.

In this case, the balancing ratio is 100%. The resource's monthly FCM payment is reduced to reflect that it under-performed its CSO by 10 MW during the scarcity conditions. Its monthly payment becomes

(FCA *Price* × CSO MW) + (–10 MW × *Performance Payment Rate*).

The first term is the base payment. If the FCA price is $3 / kW-month, the resource's base payment is $300,000 per month. If (say) the performance payment rate is $5,000 / MWh, the resource's performance payment is (–10 MW × $5,000 / MWh) = –$50,000. The resource's under-performance during the operating reserve deficiency reduces its FCM payment for the month by $50,000, to $300,000 – $50,000 = $250,000.

► **Example 3.** Same assumptions as in Example 2, but assume load and reserve requirements equal 60% of total CSO MW during the scarcity condition.

In this case, the balancing ratio is 60%. The resource's score for the event is (90 MW – 60% × 100 MW CSO) = +30 MW. Assuming the same performance payment rate as in Example 2, at $5,000 / MWh, the resource's total performance payment for the month is (30 MW × $5,000 / MWh) = $150,000. Its over-performance during the operating reserve deficiency increases its FCM payment for the month to $300,000 + $150,000 = $450,000.

---

[16] Scores are duration-weighted because scarcity conditions can occur for fractions of hours. For clarity, we assume scarcity conditions occur in hourly durations in examples that follow.

► **Example 4.** Three hours of scarcity conditions occur in a month, each when load and reserve requirements equal 60% of total CSO MW. A resource with a CSO of 100 MW performs at 90 MW, 0 MW, and 20 MW during each hour, respectively.

The resource's total score for the month is –70 MW. This is calculated as a score of (90 MW – 60% × 100 MW) = +30 MW for the first hour, (0 MW – 60% × 100 MW) = –60 MW for the second hour, and (20 MW – 60% × 100 MW) = –40 MW for the third hour; a total of 30 – 60 – 40 = – 70 MW. Assuming the same performance payment rate as in the previous examples, at $5,000 / MWh, the resource's performance payment for the month is (–70 MW × $5,000 / MWh) = –$350,000. Its total payment for the month is $300,000 – $350,000 = –$50,000. The resource's total monthly payment is negative (the resource pays for its lack of performance).

Properties and Interpretation

This pay-for-performance approach has several properties that affect incentives and risk. In this section we clarify these properties, address how the balancing ratio affects a supplier's 'upside' and 'downside' risk, and summarize locational considerations for performance incentives.

Performance Payments are Transfers Among Suppliers

During scarcity conditions, some resources are likely to over-perform and reduce the severity of reserve or energy deficiencies due to others' under-performance. In doing so, the total performance payments charged to resources that under-perform are used to compensate the resources that over-perform during the scarcity condition. Effectively, the FCM performance incentives amount to financial transfers from under-performing to over-performing capacity resources during the times when additional resources are needed to maintain system reliability.

► **Example 5.** One hour of scarcity conditions occurs in a month, when load and reserve requirements equal 60% of total CSO MW. Unit A has a CSO of 140 MW, and performs at 0 MW. Unit B and Unit C each have a CSO of 80 MW, and each performs at 80 MW. There is an operating reserve deficiency of 20 MW.

In this case, Unit A's performance payment for the month is based on its score of (0 MW – 60% × 140 MW) = –84 MW. Units B and C have a score of (80 MW – 60% × 80 MW) = +32 MW each. Assume as before a performance penalty rate of $5,000 / MWh. Then Units B and C receive a performance payment for over-performing of 32 MW × $5,000 = $160,000 each. This $320,000 total performance incentive payment is (more than) offset by the performance payment charged to under-performing unit A, which is –84 MW × $5,000 = –$420,000.

In this example, note that there is a difference of $100,000 between the performance payment debited to the under-performing supplier and the performance payment credited to the over-performing suppliers. This difference is always equal to the size of the operating reserve deficiency, 20 MW, times the performance payment rate: 20 MW × $5,000 / MWh = $100,000 for the hour. It occurs because during an operating reserve deficiency, total under-performance exceeds over-performance (if not, there would be no deficiency). The net of all performance payments could be rebated to loads, which experience a reduction in service reliability (*viz.*, the ability of the system to recover in a timely manner from contingencies) when under-performance results in scarcity conditions.

Performance Risk

A resource's performance score and the balancing ratio determine the 'upside' and 'downside' performance payment risk a supplier faces during a scarcity condition. By design, resources face less exposure to losses (negative performance payments) during low-load conditions, when many resources may be offline. They face larger maximum losses during high-load conditions, when they are more likely to be online. These properties tend to reduce a resource's downside risk.

The reason resources will tend to have a lower likelihood of incurring losses when their maximum exposure to loss is larger is because of the balancing ratio. It reduces each resource's maximum exposure to losses when load declines. In addition, the balancing ratio ensures there is always an 'upside' incentive payment potential for any resource that produces at (or above) its CSO MW during a scarcity condition. An example illustrates these properties.

► **Example 6.** A particular resource normally operates at its full 100 MW CSO when system load is 15 GW or greater, and is normally offline otherwise. Assume reserve requirements are 2 GW, and the system total CSO MW is 30 GW. If load is low at 10 GW during a scarcity condition, the balancing ratio is (10 GW + 2 GW) / 30 GW = 40%. The offline resource's maximum loss is the performance payment rate applied to 40% × 100 MW CSO = 40 MW. If load increases to 16 GW, the resource's maximum loss increases: it becomes the performance payment rate applied to 60% × 100 MW = 60 MW. However, at a load level of 16 GW, this resource is normally online. That yields an 'upside' payment for performing at its full 100 MW CSO equal to the performance payment rate applied to a score of (100 MW – 60% × 100 MW CSO) = 40 MW for the duration of the scarcity condition.

What would happen if a balancing ratio was not used in the performance score? In that situation, there would be no 'upside' reward for strong performance during scarcity conditions. The score would be the resource's Actual MW less CSO MW, which for most resources is zero or negative. In simple terms, without a balancing ratio a resource would face a penalty anytime its performance is less than its full CSO MW. With a balancing ratio, the resource is rewarded for performance that exceeds a (load-dependent) threshold, and the threshold is below its full CSO MW. In this sense, the balancing ratio ensures a supplier faces both 'upside' reward and 'downside' risk during scarcity conditions, depending on the level of energy and reserves its resources provide.

What would happen if there are no scarcity conditions at all? By design, if there are no scarcity conditions, resources are paid the FCA price without performance adjustments. This ensures that suppliers would fully recover the missing money revenue stream in the event that suppliers' performance, in the aggregate, is so strong when the system conditions are tight that operating reserve deficiencies never occur. Moreover, it provides suppliers with a degree of insurance in the event there are zero scarcity conditions in a year with mild weather and unusually few major contingencies.

Performance Payment Caps

In the performance payment formula above, there are no caps on a resource's total performance payments. This reflects an important concern over the adverse incentive consequences of explicit performance payment caps. If the potential downside from under-performance is capped in some way, it will undermine the incentives to take actions to improve the likelihood of performing when needed to ensure reliability. Specifically, if a resource were to reach a monthly or annual cap on its

under-performance charges, then its FCM revenue would no longer be reduced when it under-performs in the future. At that point, the resource no longer faces the appropriate incentives to perform that all other resources do, as even complete non-performance will not reduce its FCM revenue further.

If a resource is physically out of service for an extended duration during the capacity commitment period, the supplier should trade out of the resource's capacity supply obligation. This will also transfer the pay-for-performance incentives associated with the obligation. This transfer is efficient, as it ensures another facility has strong performance incentives to deliver the energy and reserves that the out-of-service facility can no longer provide. However, if a supplier with an extended facility outage reaches a cap on its non-performance charges, the supplier has little incentive to trade its capacity obligation to another resource that can provide replacement capacity to the system.

Locational Considerations

In implementation, this pay-for-performance approach will need to account for certain zonal and locational considerations. First, while the approach is described above in terms of system-level requirements, it should be implemented at the zonal level as well. This means that if an operating reserve deficiency occurs in a particular reserve zone, but at the time all system-level requirements are satisfied, the performance incentives would apply to resources in the zone where scarcity conditions are present.

Second, in real-time operation of the power system, it can be necessary to limit a resource's energy and reserve supply to less than its full physical capability during an operating reserve deficiency. This can occur if the full output of the resource would violate a real-time transmission operating limit. From a conceptual standpoint, incremental production by a resource has no value if the transmission system cannot deliver it to load. It would not be appropriate to reward a resource for increments of energy or reserves that cannot alleviate the scarcity condition (in fact, doing so could adversely impact reliability). Accordingly, the pay-for-performance incentives would be limited, when necessary, to the maximum actual MW value that the transmission system can accommodate from an affected resource in real-time.

Performance Payment Rate Considerations

The central design element of the pay-for-performance approach that must be set administratively is the performance payment rate. Here, several considerations come into play.

The performance payment rate, in combination with the price of energy and reserves, determines the marginal incentive to provide energy or reserves during scarcity conditions. Under the pay-for-performance approach, this marginal incentive is the same whether the balancing ratio is high or low; and it is the same whether or not a resource is operating above or below its CSO MW. Every time the system is in scarcity conditions, all capacity resources have the same incentive—at the margin—to deliver additional energy or reserves.

Economic theory offers considerable guidance regarding how high the performance payment rate should be set. Consider again Figure 1 in Section 2. In a normal, properly functioning market, the size of the incentive for the highest-marginal-cost seller to produce when demand reaches the

market's short-run capacity is the difference between the efficient price, *p**, and the marginal supplier's incremental cost (in Figure 1, supplier C at $150). Over time, this difference must cover the marginal resource's fixed costs in order for it to be willing to enter (or not to exit) the market.

A hallmark of competitive markets is that the marginal supplier makes a normal rate of return on its investment (commensurate to its capital risk), and nothing more. This means that, in expectation, the missing money component of its total revenue will be just enough to cover its total fixed costs (including return on investment). If we know the marginal supplier's (annualized) total fixed cost (*FC*) per unit capacity, and let *N* represent the number of hours per year that demand reaches the market's short-run capacity limit, then the average missing money per scarcity hour that the marginal supplier must expect to recover (per unit of capacity) if it is fully available is the ratio: *FC / N*.

What does this imply, in practical terms? It means that, on average, an efficient market would set the marginal incentive to produce during scarcity conditions at the marginal cost of energy *plus* the ratio *FC / N*. With an energy market that sets price at short-run marginal cost during scarcity conditions, the FCM will therefore provide the appropriate performance incentive if

$$\textit{Performance payment rate} = FC / N.$$

A brief translation of terms: In practice, the marginal supplier in the energy market is generally approximated by a 'proxy' benchmark peaking unit operating on expensive fuel (such as distillate oil). *FC* represents the net annualized fixed cost (including the cost of capital) that a supplier would require, in the form of capacity market revenue, to break even with this proxy unit. While estimates of this cost can vary, an educated estimate for present purposes is $105,000 per MW-year.[17] Second, *N* represents the total number of hours that scarcity conditions would occur on the power system when the system is 'at criteria', meaning total capacity exactly equals the installed capacity required. This value is difficult to calculate with simple methods, but the planning models of the New England system used to determine the ICR indicate it is approximately 21. Putting everything together, the core economic insight is that, in application, a performance payment rate on the order of $105,000 / 21 = $5,000 per MWh is needed.

In practical terms, the performance payment rate needs to be set high because resource owners must be incented to take costly actions that help avoid, and reduce the severity of, adverse reliability conditions. A performance payment rate of *FC / N* provides, by design, the minimum incentive necessary for suppliers to invest in resources they expect will enable them to be available—and fully perform—in all *N* scarcity condition hours expected annually.

Even when the market is long on capacity, as is New England at present, this performance payment rate helps suppliers make a business case for operational-related decisions that will improve system reliability. As discussed in Section 3, these decision might include installing dual-fuel capability, improved technology to better communicate dispatch signals, improved maintenance and staffing, short-notice or no-notice gas supply arrangements with pipelines, maintaining the resource in a warm state to reduce start times during tight system conditions, or any other action that makes the resource more likely to be available and able to respond to scarcity conditions on the power system. In contrast, if the performance payment rate is small relative to the costs of these actions, it is not reasonable to expect that suppliers will have the incentive to take these actions and improve their

[17] Shaw Consultants International, Inc., *Benchmark Price Model* (February 2, 2012) for ISO New England Inc.

service reliability. If the performance payment rate is low, accepting the under-performance charges might be the more economic choice.

A related point merits note. In this approach, the performance payment rate is independent of the current FCA clearing price. There is good reason for this. Conceptually, the value of avoiding a reserve deficiency is the same whether the capacity market is tight, or has excess supply. In the former case, new entry sets a relatively high FCA price, and in the latter case of excess supply, the de-list bid of an existing resource sets a low FCA price. One of the shortcomings of the FCM's existing Shortage Event penalty structure is that it fluctuates based on the FCA price. This tends to undermine incentives for resource performance and availability when there is excess capacity, and the region has observed these conditions in recent years.

Risk

As illustrated above in Example 4, there is the possibility that strong performance incentives could turn a poorly performing resource's total FCM revenue negative in a specific month, or year. That possibility is not a design objective *per se*. Rather, the intent (and logic underlying the *FC / N* formula) is to set the performance payment rate such that, in expectation, a supplier with a resource that *never* performs during any scarcity conditions would, over time, find that its base FCM payment is offset by its total under-performance charges. In this way, a resource that has a reliability value of zero to the power system would not earn the missing money revenue stream that, in an efficient market, it would not be able to earn.

Still, scarcity conditions are far from completely predictable, and their frequency and severity can differ from expectations in any given month or year. As a result, suppliers will face the risk that if there are a larger-than-expected number of scarcity conditions, a resource that significantly under-performs may experience negative monthly payments. More generally, over time every resource should expect that it will be off-line during some scarcity conditions. Even a highly efficient capacity resource that over-performs most of the time may have some months in which its net FCM revenue is less than its base FCA payment.

This risk is, to some degree, inherent to any efficient system of strong performance incentives. Without an 'upside' reward for strong performance, and a 'downside' risk of lower revenue for poor performance, desirable performance and investment cannot be achieved and reliability risks will remain. Strong performance incentives will increase the financial risk that capacity suppliers face relative to the current situation, and suppliers should be expected to account for the increased financial risk they face in their FCA bids.

One practical consequence of increasing suppliers' performance risk is that the mitigation review of FCA bids will need to account for the new risks they face. For a poorly performing resource, an expectation of net negative performance payments is a legitimate going-forward cost of acquiring a capacity supply obligation. Under the present tariff, any risk that can be analytically supported (by the participant) may be included in the going-forward costs used to evaluate its FCA de-list bid.[18] The

[18] Market Rule 1, §III.13.1.2.3.2.1.3.

existing tariff provisions governing this treatment may need to be enhanced, and it may be necessary for these provisions to provide more specific guidance.

Capacity Investment Incentives

In addition to short-term performance incentives, the pay-for-performance approach will also alter incentives for investment in capacity resources. Relative to the current situation, it enhances the incentive to invest in either (1) low-cost, highly efficient capacity resources, or (2) highly flexible, highly reliable resources. Transitioning the existing fleet of capacity resources to a system with these attributes will provide the lowest-cost means to ensure reliability over the long run.

On these points, a few examples are illustrative. Consider, at one end of the spectrum, a high-efficiency generation technology that, in the energy market, would operate at its capacity in economic merit for 95% of the hours per year. Such a resource is likely to be operating during nearly all of the scarcity conditions each year. Moreover, since some of these scarcity conditions may occur when load levels are modest or at least below ICR, by operating at its CSO during these times the unit will be consistently over-performing during scarcity conditions. It can therefore expect to receive performance incentive payments during scarcity conditions, and experience few under-performance charges. This resource contributes greatly to system reliability by being available, to the utmost limits of its capability, in all (or nearly all) scarcity conditions each year. Accordingly, the pay-for-performance approach will enhance the incentives for investment in this type of low-cost, high-availability, and highly reliable resource.

At the opposite end of this spectrum, the pay-for-performance approach provides disincentives for highly unreliable resources to accept capacity supply obligations, or to remain in the capacity market. As an illustrative example, consider an old, expensive resource used for peaking service. Because of its high marginal costs, it operates less than 5% of the hours each year. Assume this resource takes twelve or more hours of effort to start up, and even then does not start reliably. Such a resource might miss many, or possibly all, of the scarcity conditions each year. Even if a scarcity condition is anticipated many hours in advance, the emergency is likely to have passed by the time this resource could get online.

Because it is likely to be unavailable for many of the scarcity conditions each year, the actual reliability value of this resource is low. Accordingly, the appropriate reward for it in the capacity market should be close to zero: As discussed in Section 2, in an efficient market, a resource that is not available whenever short-run capacity limits bind would not receive the missing money revenue stream. Similarly, the performance charges under a pay-for-performance approach will significantly reduce this chronic under-performer's total FCM revenue.

Last, consider a generation technology or demand response resource with high operating costs, but that is highly flexible operationally. It has a low capacity factor annually, but it can deliver its entire CSO MW onto the grid in ten minutes or less, and does so consistently when called upon by the ISO. Since many scarcity conditions occur when load and reserve requirements are well below ICR, this fast-responding resource can expect to receive performance payments for over-performing during essentially all scarcity conditions. A fast-responding, flexible unit such as this contributes greatly to system reliability, as it consistently delivers energy (or reserves) to its utmost capability at precisely the times when they are needed the most to ensure reliability. Accordingly, the pay-for-performance approach will reward this resource for the value of its highly reliable service. Moreover, the pay-for-

performance design will provide incentives for investors to develop these highly flexible, highly reliable resources.

Ultimately, these capacity investment incentives will lead to an evolution in the New England resource mix that promotes low-cost, highly efficient resources and highly flexible, highly reliable resources. This will emerge, in part, through changes in the bids of capacity resources in future capacity auctions. Resources that are unreliable and costly to operate will need to submit higher offers into the FCA, based on their expectation of performing poorly and facing non-performance penalties during the commitment period. These resources will be less likely to clear the auction, relative to today's situation, and thus more likely to exit. In contrast, the rewards for strong performance that will accrue to highly reliable resources will enable them to profitably make lower offers in the FCA, and will therefore be more likely to clear future capacity auctions.

Benefits of the Pay-for-Performance Approach

This performance scoring approach has several important attributes. First, it meets the design principles stated above. It pays the aggregate amount necessary to induce sufficient capacity (as determined by the FCA), pays it for performance during scarcity conditions so as to provide economically sound incentives, and applies equally to all resources, regardless of technology.

The pay-for-performance design provides a strong performance incentive to provide energy and reserves during scarcity conditions. The incentive can be set at a consistent level that reflects an appropriate marginal incentive to produce, and for suppliers to take actions that will help reduce the frequency and severity of scarcity conditions. In effect, the pay-for-performance approach provides capacity suppliers with both risk and reward at the right times—the times when the system cannot simultaneously meet the energy and reserve requirements that are essential to reliable service.

These incentives will, over time, lead to four important benefits:

► **Operational-Related Investment.** The payment performance rate must be significantly higher than the (marginal cost) energy price during scarcity conditions. This will provide strong incentives, and the financial capability, for suppliers to make operational-related investments that ensure resources are available during scarcity conditions.

► **Increased Resource Responsiveness and Flexibility.** Resources will have the incentive to improve operating practices, pursue incremental capital investments that shorten start times and increase ramp rates, and so on. These changes will increase their performance and the financial reward for operating during scarcity conditions.

► **Cost-effective Solutions.** Uniform performance incentives enable individual suppliers to select solutions that work best for the technology and features of their specific resources. This market-based approach rewards suppliers that pursue the most cost-effective means to improve performance and availability.

► **Efficient Resource Evolution.** Finally, the incentives provided by this pay-for-performance design will, over time, lead to a change in the outage rates of the New England resource mix and directly improve resource availability. Relative to the current situation, reliable resources will be more likely to clear in the FCA, and unreliable and higher-cost resources will be less likely to clear in

the FCA. This will result in a cost-effective evolution of resources that steadily improves system reliability.

Costs of the Pay-for-Performance Approach

The consequence of increasing the risk of lower FCM revenue due to non-performance is that suppliers will take actions to reduce their non-performance risk. Those are the desirable actions noted above. However, these risk-reducing actions will have costs, and these costs will be reflected in resources' capacity bids in future capacity auctions.

It is difficult to predict the effect on capacity market prices as a result of these changes, but it is likely to result in an incremental increase in capacity costs relative to the current situation. While this qualitative impact appears likely, it may be challenging for the ISO to provide quantitative guidance on how it may impact future capacity costs. The reason is that the impact of new FCM performance incentives on future capacity clearing prices will depend on the interaction of a number of different factors, and variation in how individual suppliers respond to these incentives.

First, and most importantly, the impact on capacity clearing prices will depend on how suppliers anticipate performing during scarcity conditions. This may vary significantly from one resource to the next, depending upon its operating characteristics and energy market participation. Second, the impact on capacity prices will depend upon the incremental costs of new investments and operational practices that suppliers undertake to improve their performance. As highlighted in Section 3, there are a range of different operational-related incremental investments that suppliers may undertake, and the types of investments that are most cost-effective for one resource may differ from the solutions undertaken by other resources. These investments will interact with, and alter, suppliers' calculations of how they will perform during scarcity conditions. The cost of these investments will affect capacity clearing prices to the extent they affect the marginal resources in the capacity auction, which may be different resources in the future than today. Third, as described above, poorly performing resources will account for the greater risks they face through their capacity bids, and it is conceivable this may lead some resources to retire sooner than they would without FCM performance incentives.

Ultimately, creating FCM performance incentives that reflect efficient, competitive market principles will enable New England to maintain a reliable power system at the lowest long-run cost. Deviation from these principles will come at a greater cost, both in terms of increased reliability risk and the need to purchase additional services and resources to compensate for resources that do not perform well. Paying resources that are not performing during scarcity conditions will inappropriately reward, and provide incentives for, resources that do not contribute to the system's reliability needs. This would ultimately require the acquisition of additional resources, at additional expense.

5. Continuing Efforts

The ISO intends to pursue, through the stakeholder process, long-term changes to the FCM to enhance resource performance and availability. We recognize that these changes will require a significant amount of time and effort from the region, but they are necessary to ensure the reliability of the power system and the competitiveness of the market structure that the region has adopted. We look forward to the opportunity to discuss these issues with stakeholders.