A proposal for the structure of a capacity market for a competitive wholesale electricity market: Advance funding for the right and obligation to provide capacity.

Revised October 31, 2002

Tom Welch, Chairman of the Maine Public Utilities Commission

Introduction

This proposal continues to be a work in progress. The basic principles remain the same; the changes attempt to deal with some of the practical and theoretical difficulties in earlier versions.

The FERC SMD order acknowledges the need for some form of capacity assurance "overlay" on the competitive wholesale market. I agree with FERC's conclusion in that regard, but do not believe that the FERC proposal in its details is consistent either with a competitive retail market (because it seems to assume that LSEs are stable entities), and because of the FERC insistence upon linking capacity assurance to physical assets makes entry problematic and the participation of DSM and transmission-based solutions virtually impossible. What follows is the outline of an alternative to the FERC approach that should, in my view, ensure the politically necessary level of capacity adequacy to persist with the least possible interference (which is not to say no interference) with the energy and related markets.

In summary, the FERC should require a capacity market that ensures that the dollars that are collected from market participants go into the pockets of people who are subject to an enforceable obligation to provide deliverable energy when and where it is needed at an energy price that is not itself the product of market power or scarcity. The RTO, with appropriate input from market participants and representatives of the public interest, and oversight by FERC, should be responsible for estimating future capacity needs and operating the market designed to ensure that those needs are met.

I believe that the claim by some economists that an unconstrained energy market alone could provide adequate incentives to ensure sufficient capacity to sustain a workably competitive electricity market may be theoretically sound but, in my view, it is politically and practically undesireable. The problem of whether there is adequate capacity is not just an economic question; it is just as fundamentally a political question. Whether a smoother price and supply curve produces a better long term allocation of resources or not, the public will not, and should not, tolerate a situation in which the lights go out periodically, or prices rise to crippling levels, in the name of "creating appropriate economic incentives for new generation." Moreover, the same economists who support an unconstrained energy market recognize, but suggest the public tolerate, a level of market power (as a way to ensure recovery of capital costs by generators that do not run very often) that I find utterly unacceptable: the public should not be at the mercy, during hours of capacity shortage, of the whims of bidders into a market where every bid must be accepted. The unavoidable conclusion is that, for at least the foreseeable future, there is a governmental responsibility to *ensure* to the extent possible, and not just assume, that adequate

capacity for both reliability and effective competition will exist at all moments, and not just on average.

It may be the case that, once demand elasticity reaches the point where the demand and supply curves are equally flexible, regulatory intervention into the market (in the form of planning and capacity assurance) will become unnecessary. In the longer term, customers must have opportunities for effective demand response; no genuinely competitive market can be sustained if one side of the supply/demand equation is fixed. (Indeed, as the recent FERC cost/benefit study of RTOs suggests, demand response can achieve very substantial cost savings regardless of what structure the markets take.) At least in the near term, however, some form of capacity market is required to help assure the public that there will be adequate electric capacity to provide reliable service and prevent the exercise of market power.

Any mechanism to ensure adequate capacity should meet at least four objectives. It should interfere as little as possible in the competitive market. It should ensure that we achieve an acceptable level of reliability at a reasonable cost. It should create a level playing field in which supply and demand side contributions toward maintaining reliability are equally valued and encouraged. And it should be flexible enough to allow for modification and, at least in principle, elimination. The current ICAP markets fail these tests miserably.

In today's ICAP markets, the money goes to people who, because they own existing generation, have every incentive to create *shortages* of capacity, rather than to firms that will build the surpluses needed to sustain a competitive energy market. Most ICAP recipients, particularly those who own substantial amounts of existing generation, understand that if they build new plant, the effect will be to reduce prices and revenues in both the ICAP and the energy markets from their existing portfolio. To put it bluntly, the current ICAP markets appear to operate as a mechanism to transfer wealth from load to generation in the hope, though without any reasonable expectation, that those receiving the wealth will act contrary to their own self-interest and ensure a sufficient future surplus of capacity to dampen energy prices. Clearly a system that connects rewards with benefits more directly (to say nothing of more logically) is required.

The proposal

There is an alternative to the current ICAP approach, and to the unrealistic alternative of eliminating capacity obligations entirely. The proposal shares some characteristics of the FERC SMD capacity adequacy proposal, in that both look a few years into the future to ensure adequacy, but also differ in important respects. The proposal I favor would involve the payment, by load, of an amount sufficient to ensure that sufficient capacity will be available into the intermediate term future. The amount of capacity needed would be determined by the RTO; the money would be collected from current load serving market participants; and the money would be paid "on delivery," i.e. on "delivery" in the year for which the capacity is promised. Importantly, and differently from the FERC approach, there is no direct link between any particular load and any particular capacity. The reason for uncoupling particular load from particular capacity is that, in my view, the capacity requirements of the system are just that: i.e. requirements of the system as a whole, not of any particular participant. While, in the past, it

may have made sense for a vertically integrated utility to ensure, for its own customers, adequate capacity, in a competitive retail market, where a capacity shortage (however created) results in higher prices for all load (through the energy market), continuing to link load to capacity is both impractical and illogical. Sufficient capacity to minimize market power and ensure reliability for the system as a whole is an obligation of the whole, in exactly the same sense as the cost of operating the central dispatch of the system is an obligation shared by all.

The details of the proposal, and a discussion of the reasons for each element, are set forth below.

Step 1: Estimation of Future Capacity Needs

The RTO (or equivalent) would, each year, with appropriate market participant and public (including regulatory) input, develop projections of "need" a "target year." The target year should be far enough in the future to permit, for the area in question, time for planning and construction of the solution bid into the capacity market, but no longer, because the further out the planning process looks, the greater the risk of substantial estimation errors and the likelihood of the recreation of stranded costs in the form of unneeded capacity. I suggest that a three or four year period might achieve the correct balance. The need for capacity would be subdivided into various categories, such as energy (or demand reduction) available on 10 minutes' notice; energy available on an intermediate period notice; and energy available at a high capacity factor (these needs could, but need not be, satisfied by peaking, intermediate, and base load generation respectively). The RTO would also identify any zones that had needs beyond those for the system as a whole (where, for example, existing transmission could not bring all the energy needed into the zone). If there is sufficient concern about supply diversity, bids could be done by plant type, with reservations for those run by fuels other than gas (or whatever else seemed to threaten diversity). In making the estimates of future capacity needs, the RTO would solicit comment from all market participants and public interest entities, and the RTO decision would be subject to review and approval by the FERC (or a regional regulatory body, should one be established).

The RTO would calculate the total demand needed at a level not only sufficient to achieve an appropriate reserve margin for reliability, but also to achieve the margin needed to ensure that the wholesale market remained workably competitive throughout the target year. This might, or might not, require some level above the level needed for reliable operation.

Step 2: Bidding and the nature of the obligation

In the "bid year" (for a 2002 bid year, in this example, the target year would be 2006 if a four year planning horizon is chosen), the RTO would hold an auction to award certificates for target year. The obligation imposed on a winning bidder would be to deliver, physically, the amount of energy (or demand reduction) successfully bid at the location specified in the auction (in the case of a system-wide auction, to the PTF or equivalent of the system; in the case of a zonal auction, into the PTF within the zone), and to do so bidding into the energy market at a price ("strike price") specified in the auction. The strike price should be set at a level that, in the

overall context of the political and economic situation, is sufficiently low to avoid catastrophic spikes and to provide some protection against the unconstrained exercise of market power but sufficiently high to allow a reasonable degree of volatility (the latter in part to ensure that the opportunities for DSM and hedging are not eliminated). The range of reasonable strike prices probably lies between $1000 and $200/MWh. The energy (or reserves) would be subject to recall by the RTO.

The bid would be in the form of a request for payment to the bidder of a specified dollar payment per MW of deliverable energy (or demand reduction). No payment would be made to any successful bidder pursuant to the certificate unless the RTO certified the bidder in the target year as providing the promised product. Each winning bidder in the capacity auction would receive (or pay) the amount bid (rather than a "clearing price"). Certificates could be traded once issued, so long as the new certificate holder agreed (and could be bound) to its terms. In the energy market (during the target year), certificate holders would receive the clearing price, and their bids would be constrained only when "called" by the RTO. When "called," the bids would be capped at the strike price (though lower bids would be permitted); at all other times, certificate holders could bid under the same constraints as all other market participants.

Because the full price of a new unit (or other solution) might result in a very high bid if it could be collected only for one year, it might be necessary to allow bids (or, indeed, require bids) that had both a multi-year obligation and a multi-year payment. Further analysis is needed on this point.

Step 3: Collecting and Disbursing the Money

Once certificates were awarded, cost of payment would be collected by the RTO from the load serving entities based on their proportionate load either in the bid year or in the target year. Each approach has advantages. Collecting in the bid year would probably increase the ease of financing projects. Once collected, the money would be held in escrow and distributed to the certificate holders who perform their obligations in the target year. The use of escrow should allow those who need financing to secure it (since performance on the certificate would give the developer, and the banks, a right to the cash). Collecting in the target year, on the other hand, avoids having a large pot of money sitting in the RTO coffers, and also matches the payment better with the customers who benefit from the capacity. Collecting from load as a whole reflects the principle that the capacity payments are not intended as a link between any particular LSE and any particular resource; they are, instead, payments made by the consumers in the market as a whole to ensure that future consumes will continue to enjoy adequate capacity and robust competition. This is not, in my view, a situation in which there are "free riders;" the analogy is more that everyone is riding in the same boat, and everyone has an equal interest in it remaining afloat

Virtues of the Proposal

The model proposed here has the virtues of relatively small estimation risk (because the planning horizon is limited), looks far enough ahead to ensure that there is time to actually build the needed facilities, provides the security of a future source of cash to allow financing for those plants; gets the money to the people who will be providing the capacity when and where needed; and, not least, provides a structure, under the supervision of the FERC, that will assure the public that regulators and RTOs are actively ensuring that the lights will stay on, and prices will reflect competition and not avoidable shortages, both today and into the future.

There would be, without doubt, a significant impact on the energy market as a result of this proposal. For one thing, bidders in the energy market would no longer have to consider whether the gap between operating costs and clearing price would be sufficient to cover capital costs: those costs could be recovered in the capacity auction. Moreover, the strike price obligation would, as a practical matter, likely act as a cap on prices. Most significantly, ensuring adequate capacity to reduce opportunities for the exercise of market power and dampen volitility would likely result in a smoother price curve. These are not trivial costs. On the other hand, failure to ensure sufficient capacity is likely to lead to more market power (with an uncertain but probably not trivial cost) and, if reliability is threatened, a backlash against the development of markets as a whole.

Further Steps

This and any other capacity assurance model need to be reviewed and tested to ensure that there are no opportunities for "gaming." Moreover, further iterations should (and will) include a "follow the money" analysis. Particular issues that clearly need discussion include whether the "pay as bid" structure, combined with the opportunity for demand and transmission (in combination with generation) to participate in the auction is sufficient to eliminate market power (defined as something held by any bidder who knows that his bid must be accepted); whether supplemental auctions might be required if there amount under certificate appeared to be falling short of the actual needs in the target year; whether any kind of "progress" obligation should be imposed, and as a supplement or alternative what the penalties (beyond non-payment) for failure to perform should be, and how collection of those penalties could be enforced.