UNITED STATES OF AMERICA
BEFORE THE
FEDERAL ENERGY REGULATORY COMMISSION

Remedying Undue Discrimination	)	
through Open Access Transmission Service	)	
and Standard Market Design	)	Docket No. RM01-12-000

AFFIDAVIT OF FRANK A. WOLAK
ON BEHALF OF
THE ELECTRICITY CONSUMER RESOURCE COUNCIL,
THE TRANSMISSION DEPENDENT UTILITY SYSTEMS,
BUCKEYE POWER, INC.,
GREAT RIVER ENERGY,
WOLVERINE POWER SUPPLY COOPERATIVE, INC., AND
EAST TEXAS ELECTRIC COOPERATIVE, INC.

City of Stanford, California	)	
County of Santa Clara	)	ss:

1. Introduction

I, Frank A. Wolak, being duly sworn, depose and state as follows: I am a Professor of Economics at Stanford University. I began my work on energy and environmental issues at the Los Alamos National Laboratory (LANL) in 1980. The following year I entered graduate school at Harvard University, where I received an S.M. in Applied Mathematics and Ph.D in Economics. For the past fifteen years, I have been engaged in a research program studying privatization, competition, and regulation in network industries such as electricity and natural gas. A major focus of my academic research is the empirical analysis of market power and, more generally, market design issues in newly restructured electricity markets. I have studied the design and operation of the PJM (The Pennsylvania, New Jersey, and Maryland Interconnection), New York, New England and California electricity markets, as well as virtually all restructured electricity markets currently operating around the world. Since April 1, 1998, I have been the Chairman of the Market Surveillance Committee (MSC) for the Independent System Operator (ISO) of California's electricity supply industry.

I have been asked by the members of the Coalition sponsoring this affidavit to propose methods for local market power mitigation under the Commission's Standard Market Design (SMD) that achieve the goals laid out in the "Remedying Undue Discrimination through Open Access Transmission Service and Standard Electricity Market Design Notice of Proposed Rulemaking" (SMD NOPR) issued by the Commission on July 31, 2002.

I would like to emphasize that a crucial pre-condition necessary for the success of the local market power mitigation methods described below is an overall market structure that will support a workably competitive wholesale electricity market. Specifically, the amount of generation capacity necessary to serve demand in this market during the vast majority of hours of the years must be owned by a substantial number of independent firms, there must be no significant legal impediments to the entry of new suppliers, and the transmission network must not substantially limit the quantity of electricity that can be economically transferred across locations in the market or into the market from neighboring regions. Moreover, the retail market structure supporting this wholesale market should provide ample opportunities and strong incentives for active participation in the hourly wholesale market by a substantial fraction of final consumers, either directly (if there is direct access) or through the intermediary of a load-serving entity (LSE). Unless these pre-conditions for a workably competitive wholesale market exist, implementing the local market mitigation methods described below will do little to protect consumers from significant financial harm, because the underlying market structure cannot support a workably competitive wholesale market.

I would also like to emphasize that my comments focus on local market power mitigation in the context of the Regional Transmission Organization (RTO) and/or Independent Transmission Provider (ITP) market structure with a spot wholesale electricity market that uses locational marginal pricing (LMP) like the PJM market. I will first demonstrate why local market power mitigation is necessary in all existing and proposed RTOs/ITPs with spot wholesale electricity markets in the United States (U.S.) for at least the next decade, whether or not they use LMP. This conclusion is the result of the following logic. Historically, transmission networks serving the vast majority of consumers in the U.S. were designed for a market structure where a combination of local generation and transmission lines bringing distant generation all owned by the same entity were used to meet an annual energy need for a local geographic area. The dominant industry structure in terms of number of customers served was the vertically-integrated investor-owned utility as the single supplier for a given geographic area. Municipal and cooperative utilities supplied electricity to the remainder of the U.S., with cooperatives providing service primarily in more sparsely populated areas.

The new ITP regime that the Commission is now attempting to implement is therefore very different from the dominant historical industry structure, where a single firm owned and operated both the transmission network and virtually all of the generation units in its control area. Under the new wholesale market regime, the operator of the transmission network is financially separate from all generation owners. Under the prior industry structure, state public utilities commissions fixed the retail price paid to the vertically integrated investor-owned utility for all of the energy it sold. In contrast, under the new wholesale market regime, competitive forces determine the prices power producers receive for their electricity.

Because of the vertically integrated investor-owned utilities' prior transmission and generation construction decisions, in many areas of the U.S., the existing transmission network may have inadequate transfer capacity to face every generation unit owner with enough competition from distant generation unit owners to elicit competitive behavior from the firm at each location in the transmission network that it serves for the vast majority of hours of the year. Instead, a firm owning a substantial fraction of the generation capacity in and around a geographic area with inadequate transmission capacity to serve local demand has an incentive to withhold energy from the market--either by bidding very high prices or by refusing to operate some of its units--to increase the prices paid for the energy it does supply.

The combination of a transmission network designed for an industry structure that no longer exists and the resulting perverse incentives for profit-maximizing behavior by dominant generators in a wholesale market regime like that described above implies that at many locations in the transmission network of all existing and proposed RTOs/ITPs in the U.S., there are a substantial number of hours of the year when only one firm or a small number of firms can meet a local energy need. Without a local market power mitigation mechanism in place at the RTO/ITP level, there is almost no limit to the price that these firms can bid to supply energy. This logic also implies that there is virtually no limit on the price that consumers located in this area would have to pay for electricity because they happen to live in an area that the former vertically integrated investor-owned utility found least-cost to serve with a combination of local generation and transmission capacity.

In any transition to such a new wholesale market regime, consumers of electricity should not be punished or rewarded for their location in a transmission network built to serve an industry structure that no longer exists. Consequently, any local market power mitigation mechanism adopted as part of a wholesale electricity market using LMP as recommended in the Commission's SMD NOPR, should ensure that all participants in the RTO/ITP share in the benefits of wholesale competition. The mechanism I propose is designed to achieve this goal. Specifically, its primary goal is to minimize the financial burden that any market participant must bear as a result of the initial conditions in the RTO's/ITP's transmission network—conditions that already exist and that are therefore unrelated to the formation of a wholesale electricity market. Although this mechanism strives to eliminate the harm to consumers from initial conditions in the transmission network, going forward it still sends the market-determined locational price signals to all market participants, so that consumers will make the appropriate electricity consumption decisions and producers the appropriate investment decisions.

This mechanism limits the incentives and ability firms have to exercise local market power in both the RTO's/ITP's spot and forward congestion management markets. Mitigation occurs in the spot congestion management market if the RTO/ITP determines that only a very small number of firms are able to provide a given local energy need. Under these system conditions, the RTO/ITP mitigates the price bids submitted by these firms only for the necessary quantity of energy. These mitigated price/quantity bids do not enter the RTO's/ITP's LMP process. In contrast, all price/quantity bids associated with the capacity from these units that is not needed to meet this local energy requirement do enter the RTO's LMP process. Several three-node DC-load-flow transmission network examples that illustrate how this mechanism would function are presented below.

Under my proposal, mitigation occurs in the forward congestion management market in the following two ways. First, the RTO/ITP is required to allocate all Congestion Revenue Rights (CRRs) to load-serving entities (LSEs) rather than auction them off to the highest bidder. Second, LSEs that own generation or a long-term contract to the output from units in close geographic proximity to the load they serve will have their CRR allocation adjusted to account for the degree to which this local generation ownership or long-term energy contract provides an equivalent physical hedge against the congestion charges these firms otherwise would pay to serve their load. As discussed more fully later, it is important to emphasize that under no circumstances would this process result in a one for one MW CRR reduction per MW of local generation ownership/control. This two-step mechanism limits the ability and incentive generation unit owners have to cause congestion and thereby increase the revenues they earn from their CRRs. I discuss the local market power mitigation properties of this allocation scheme in detail below.

The final topic of my remarks is the need for and role of an independent market monitor in an RTO/ITP. Because it is impossible to anticipate, before the start of any market, all of the instances when a firm might be able to exercise a substantial amount of local market power, it is necessary to monitor continuously the behavior of market participants. Moreover, because the RTO/ITP is a monopoly provider of system operation services for a given geographic area, it may sometimes take actions that unnecessarily impose significant costs on certain market participants. This potential conflict between the goals of the RTO/ITP management and the goal of efficient operation of the RTO's/ITP's markets implies a fundamental need for an independent market monitor or market monitoring committee. For this reason, this market monitor should be completely independent from the RTO/ITP, and one of its roles should be monitoring the performance of the RTO/ITP. My comments will first describe why I have in my own past experience found it useful to have both internal and external entities to ensure effective market monitoring. This discussion identifies potential sources of conflict between these two entities that could lead to inadequacies in the RTO/ITP's market monitoring process. It also forms the basis for my subsequent recommendations about the optimal relationship between two such monitoring entities and the RTO/ITP management and board of directors, and the sorts of data these two market monitoring entities should collect and share among themselves to best enhance the efficiency the RTO/ITP's market and system operations.

2. Origins of Local Market Power Problem

Local market problems arise in the new unbundled wholesale market regime because the existing transmission network in virtually all parts of the U.S. is very poorly suited to support the geographic extent and magnitude of electricity trading required for a workably competitive wholesale market. Because the existing transmission network in the U.S. was largely built over the past 50 years, a period dominated by vertically integrated investor-owned utilities, it is easy to understand why this is the case. Moreover, over the past ten years, as the vertically-integrated utility regime has given way to increased wholesale competition, there has been a significant decline in investment in transmission expansion and upgrades throughout the U.S. relative to the growth that occurred during previous decades.[1]

The capacity and configuration of the transmission network and geographic location and composition of generating units throughout the U.S. were designed to be operated by vertically integrated utilities. In most parts of the U.S., these utilities are investor-owned, but there are notable government and cooperatively-owned exceptions. The key feature of all of these networks is that they were designed to take advantage of the fact that the same entity owned and operated the transmission and distribution network, as well as the vast majority of generating units needed to meet the utility's load obligations. The interconnections between the transmission networks of the neighboring geographic vertically integrated companies were primarily constructed to guarantee the engineering reliability of each control area. They were not designed to facilitate a substantial amount of across-control-area trade of electricity. They were built to ensure that if an unexpectedly large amount of generation in one control area suddenly failed, there was enough transmission capacity so that enough energy from the neighboring control area could be imported to keep the system in balance until this situation could be remedied.

[1] Hirst, Eric, "Expanding U.S. Transmission Capacity," July 2000, available from http://www.ehirst.com, provides a detailed discussion of trends in transmission investment in the US over the past 30 years.

Particularly, around large population centers and in geographically remote areas, the vertically integrated utility used a mix of local generation units and transmission capacity to meet the annual demand for electricity in the region. Typically, this utility supplied the region's baseload energy needs from distant inexpensive units using high-voltage transmission lines. It used expensive generating units located near its load centers to meet periodic demand peaks in the area throughout the year. This combination of local generation and transmission capacity to deliver distant generation was the least-cost strategy for serving the utility's load in the former regime. State retail rate regulation was then used to set the prices paid by all customers. These prices were set to allow the firm an opportunity to recover its production costs, including a return on capital.

The transmission network that resulted from this strategy by the vertically integrated utility creates local market power problems in the new wholesale market regime. The amount of transmission capacity necessary for reliable grid operation in the vertically integrated regime is significantly less than the amount necessary to face every generation unit owner with sufficient competition to cause them to bid close to their minimum marginal cost of supply in the wholesale market regime. The geographic configuration of the transmission network necessary for a competitive wholesale electricity market is also substantially different from the one best suited to the vertically integrated supplier regime. The vertically integrated regime allowed the utility to capture any potential economies of scale between among generation, transmission and distribution in serving end use demand, whereas the new wholesale market regime does not.

Under the new wholesale market regime, the owner of the local generating units may not own and certainly does not operate the transmission network. The owner of this local generation unit may not even be the LSE for that geographic area. Under the vertically integrated regime, the geographic monopolist had an obligation to serve all demand at the regulated retail price. Under the wholesale market regime, unless the owner of this local generating unit is subject to state-level regulation of its sales, the firm now earns higher profits by selling all output from the unit (that has not been pre-sold under a long-term forward contract) at the highest possible price in the wholesale market. This price depends on the bids this firm submits to supply energy. Consequently, during the periods when this firm knows that output from its units is needed to meet local demand, it is profit-maximizing for the unit owner to bid whatever the market will bear for any energy supplied to the wholesale spot market from these units.

Under the wholesale market regime, there must be enough transmission capacity into the region served by each generation unit so firms with units outside of this area have the potential to capture a large fraction of this local demand. Otherwise the unit owner may find it profit-maximizing to take advantage of the limited transmission capacity into the region by bidding substantially in excess of its minimum marginal cost of supply and only selling a small amount of energy, but at a very high price.

To provide a concrete example of a local market power problem, consider the 3-node DC-load-flow example shown in Figure 1.[2] All 2637 MW of load is concentrated at node 3 and its demand is inelastic. There is generation located at each of the nodes in the

[2] This 3-node network model is a slight modification of the one used by Cardell, Judith D., Hitt, Carrie C., and Hogan, William W. (1997), "Market Power and Strategic Interaction in Electricity Networks," *Resource and Energy Economics*, 19, 109-117.

network with marginal cost functions given in Figure 1. Each transmission link is the same length and is therefore assumed to have the same resistance. However, the link between nodes 2 and 3 has a maximum capacity of 600 MW. The other two links have sufficient capacity so that only the 600 MW link between nodes 2 and 3 is ever binding. Because this example ignores the impact of line losses on energy flows, a feasible set of energy schedules for the generators located at the three nodes satisfies the following two constraints: (1) the total amount of energy supplied by each generator equals the total demand at node 3 and (2) the flow on the link between nodes 2 and 3 is less than or equal 600 MW. I employ a 3-node example because it is the simplest transmission network model that can illustrate the local market power problem and account for the impact of loop flow on the locational marginal prices of energy.

Because of the configuration of the transmission network and the geographic location of demand, the generator located at node 3 must supply some energy or the first constraint necessary for a feasible schedule will be violated. Because of Kirchoff's Laws governing power flows, a 1 MW injection of energy at node 1 will lead to 1/3 MW flow from node 1 to node 2 and node 2 to node 3, and a 2/3 MW flow from node 1 to node 3. Similarly, a 1 MW injection at node 2 will lead to a 1/3 MW flow from node 2 to node 1 and node 1 to node 3 and 2/3 MW flow from node 2 to node 3. Consequently, defining q_i as the amount of energy injected at node i, we can write the transmission constraint on flows between node 2 and 3 as $1/3\ q_1 + 2/3\ q_2 \leq 600$. Consequently, any values q_1, q_2 and q_3 that sum to 2,637 and satisfy this inequality constraint represent a feasible set of generation schedules. Multiplying both sides of this inequality by 3 yields the expression, $q_1 + 2\ q_2 \leq 1800$. Written in this form, we can easily see that the constraint on the transmission capacity between nodes 2 and 3 implies that the generator at node 3 must produce a substantial amount of energy or the constraint $q_1 + q_2 + q_3 = 2,637$ will be violated. By inspection we can see that the minimum possible amount energy that must be supplied from node 3 is 2,637 – 1,800 = 837. This occurs when the generation at node 1 produces 1,800 and the generators at node 2 produce nothing. Consequently, the generator at node 3 is a local monopolist for at least 837 MW under this transmission network configuration and demand conditions. The generator at node 3 is pivotal for 837 MW, because regardless of the price it bids to supply this amount of capacity, physical constraints on the capacity of the transmission network imply that this quantity of energy from the generator at node 3 must be accepted or the constraint that supply equals demand at node 3 will be violated. It is important to emphasize that one implication of a 3-node network model is that if the generator at node 2 supplies any amount of energy, the generator at node 3 becomes pivotal for a quantity of energy larger than 837.

Under the former vertically integrated regime, this need to run local generation did not create a local market power problem. The vertically integrated utility simply found it least cost to serve this load with a mix of expensive local generation capacity and transmission capacity bringing in distant generation. However, in the new wholesale market regime, if a single merchant electricity supplier owns all of the capacity at node 3, this creates a local market power problem that can have both economic and reliability consequences. Any time the local demand for energy is greater than 1,800 MW, this firm is a monopolist for the local demand in excess of the available transmission capacity into the region.

How some state PUCs have decided to sell off the generating assets of the former vertically integrated utilities that they regulate has further compounded this local market power problem. In particular, these divested generating units are often purchased by the new entrants in sets of units clustered in discrete geographic areas. This implies that in many markets, all of the generating units needed to meet the remaining local demand are owned by a single firm. While there may be some cost advantages to a firm from having all of its units located close to one another, the ability to exercise local market power appears to be a major factor in the decision of these firms to purchase bundles of generating units in close proximity to one another.

As the example in Figure 1 makes clear, whether a firm is a local monopolist depends on the level of local demand, the amount of available transmission capacity into the region, and concentration of ownership of local generation. The congestion management scheme used by the RTO/ITP does not determine the extent to which there is a local monopoly problem. Figure 1 also shows that the behavior of all firms owning generating units determines whether the firm owning local generation possesses significant local market power and the magnitude of local market power that this firm can exercise. Particularly in a looped (as opposed to radial) transmission network, how a firm that owns multiple generating units operates its units in one geographic area can impact the amount of available transmission capacity to serve load in a geographic area where it owns other generating units. For example, in Figure 1, suppose that the same firm that owns the capacity at node 3 also owns the capacity at node 2. As the above discussion illustrates, the more energy the generation units at node 2 supply, the larger is the amount of energy that must be supplied by the generation units at node 3. If local demand at node 3 is sufficiently high, this firm could find it extremely profitable to increase production from its units at node 2 to reduce the amount of available transmission capacity that the generation units at node 1 can use to sell at node 3, to be a local monopolist over a larger quantity of energy.

There are several reasons why local market power problems are likely to be more severe and last longer in the U.S. relative to other industrialized countries such as the United Kingdom (U.K.). The National Grid Company (NGC) in the U.K. owns and operates the entire U.K. transmission network and real-time balancing market on a for-profit basis.[3] In contrast, under the U.S. model, most of the former vertically integrated utilities have retained ownership of their portion of the transmission network that each ISO now operates. Because of the divergent interests of transmission network owners and the separation between ownership and operation of the transmission network in U.S. wholesale market, determining and implementing transmission upgrades that are beneficial to market efficiency and system reliability is likely to be more difficult to implement in the U.S. than in the U.K. Even if upgrades with positive net benefits could be identified, the divergent interests of the transmission owners in the U.S., many of which also own local generation units, will make it more difficult to get these upgrades approved and constructed. For this reason, any inadequacies in the transmission network

[3] As a result, the U.K. regulator can provide strong incentives for least-cost market and grid operation, including congestion management costs. For example, during the early years of the market under a cost-of-service regulatory regime, total grid operations charges by the NGC showed a steady increase, going from £112.1 million in the 1990/91 fiscal year to £157.7 million in the 1993/94 fiscal year. As a result of this trend, in 1994 the Office of Electricity Regulation (Offer) instituted the Transmission Services Scheme. This arrangement allows NGC to keep cost savings or pay costs incurred beyond a target cost amount, thus encouraging NGC to minimize "avoidable costs" in managing the transmission grid. Following the implementation of this incentive regulatory scheme, grid operations costs steadily fell, despite the fact that annual total system load has continued to grow. By the 1997/98 fiscal year, the total cost of ancillary services was back down to £117.5 million.

that create opportunities for firms to exercise local market power are likely to persist for longer in the U.S. than in the U.K.

The example in Figure 1 assumes that local demand is completely insensitive to the hourly price of electricity. This is a realistic assumption for U.S. markets, at least for the next decade, because there are very few consumers purchasing according to the hourly spot price of electricity in any of the currently operating U.S. wholesale markets. Over the long-term, price-responsive end use demand can limit substantially the opportunities generating unit owners have to exercise local market power. However, regulators in many states have shown little interest in exposing retail customers to the sort of hourly price signals necessary for active demand-side participation in the wholesale electricity market. In addition, very few end use consumers currently have the interval metering technology necessary to realize the full benefits of responding to hourly wholesale price signals. Finally, for active demand side participation to effectively limit market power in the wholesale markets, nearly all customer classes must first fundamentally change how they purchase and consume electricity. The time necessary to change the attitudes of consumers and regulators and equip a large enough number of final consumers with interval meters is long enough for price responsive final demand to be at best a medium-term solution to the local market power problem. This does not mean that these demand-side changes should not be adopted, only that it is unrealistic to think that it will not take a substantial amount of time to do so.

All of these factors make it essential that the Commission have, as part of its SMD, a transparent mechanism for an RTO/ITP to mitigate local market power. This mechanism would mitigate local market power without distorting the prices paid to generating units that do not possess local market power. It should also provide the strongest possible incentives for all generation unit owners to bid as close as possible to their minimum marginal cost of supply during all system conditions.

Before describing my proposed market power mitigation measure, I will introduce an analogy with geographically dispersed markets that is useful for illustrating the cause and solution to the local market power problem. Before the U.S. interstate highway system became ubiquitous, transporting goods between U.S. cities was considerably more expensive, and in many case prohibitively expensive. Consequently, each city had to produce locally a large fraction of the goods it consumed, because of the high cost of importing goods from distant locations. Under these circumstances, local firms could often exercise significant market power through prices they charged to local consumers. The cost of entry was sufficiently high relative to the potential profits that a new supplier could expect to earn because of transportation costs into the area and the limited revenue potential of the small local market that such market power was sustainable over time.

As the demand for goods in certain areas grew, expanding the capacity of the transportation links between these areas became economic. These enhanced transportation links, in turn, limited the ability of producers in these regions to exercise local market power, because they now faced significant competition from distant producers. Moreover, the growing size of these markets implied significantly greater potential revenues from entry, particularly in the largest and fastest growing areas.

The self-reinforcing mechanism described above also implies that regions with little economic activity or prospects for growth will continue to face significant local market power problems. Unless the local market problem is extreme, it makes very little economic sense to invest in significant new transportation capacity into a small locality with little prospect of significant growth. Consequently, these local market power problems could persist for the foreseeable future.

This analogy and the self-reinforcing mechanism of growth in local economic activity providing the economic justification for expanding the transportation infrastructure between these areas, have important implications for electricity networks. The pre-interstate highway system is analogous to the vertically integrated utility regime when there was little electricity trading across control areas, because these transmission networks were designed to serve the utility's service area, not to facilitate trade. Any local market power problems associated with these transmission networks were solved by state-level, cost-based regulation of the retail prices of electricity sold by these vertically integrated utilities.

Throughout the U.S., the legacy of the vertically integrated utility regime is a transmission network that provides significant opportunities for firms to exercise local market power. The least cost network design and geographic location of generation units used by a former vertically integrated utility now creates system conditions where certain wholesale power producers face insufficient competition from distant generation to cause their expected profit-maximizing generation unit bid curve to be very close to the unit's marginal cost curve.

Because of these initial conditions in transmission networks throughout the U.S., even for large load centers or load centers expected to experience significant growth, a local market power mitigation mechanism is necessary to protect end use customers during the transition period during which new transmission capacity is built to provide all generation units in these areas with sufficient competition from distant generation units to cause these firms to bid close to their minimum variable cost of production from their units.

These initial conditions in the transmission network create a longer-term problem for smaller load centers not expected to experience significant load growth. These regions may find that a local market power mitigation mechanism is the least-cost solution to this problem in the medium or long-term. If for next 10 to 15 years, local demand can be met with the combination of the output of the local monopoly or duopoly generation owners and the existing transmission capacity into the region, it does not make economic sense to invest in a substantial transmission upgrade in order to face these local generation unit owners with sufficient competition from distant generation to cause them to bid aggressively. Much lower cost solutions are available.

Returning to the analogy with geographically dispersed markets, the transmission upgrade solution to the local market power problem for small low-growth areas is the same as building multiple highways into a remote, sparsely populated area not expected to experience much growth in order to limit the local market power of firms in this area. For the same reason that this solution makes little economic sense as public policy for constructing highways, the transmission upgrade solution would be a misallocation of scarce dollars for investments in transmission capacity. Instead, similar to the vertically integrated monopoly regime when explicit regulatory intervention was used to protect end use customers from local market power, such customers could rely on the local market power mitigation mechanism to protect them for as long as the economic drivers of electricity consumption in this area remained relatively constant.

3. The Goals of Local Market Power Mitigation under LMP

The Commission states that its SMD NOPR is designed "to harness the benefits of competitive markets for the nation's electric energy consumers, in order to meet [its] statutory responsibility to assure adequate and reliable supplies of electric energy at a just and reasonable price." For the reasons described in the previous section, a local market power mitigation mechanism that satisfies the following three criteria is essential to achieving this goal.

First, the local market power mitigation mechanism must recognize that initial conditions in the transmission network of a wholesale market using LMP can lead to significant harm to end use customers through extremely high electricity prices at certain locations in the transmission network. As stated in the previous section, these prices occur solely because of decisions by the former vertically integrated monopolist and approved by state PUCs about how to construct a transmission network and the geographic mix of generation to most efficiently meet its system-wide load obligations. Consequently, to benefit as many of the nation's electricity customers as possible, the SMD local market power mitigation mechanism must limit the harm to customers caused by a transmission network and portfolio generation units designed for an industry structure that no longer exists.

Second, the SMD local market power mitigation measures must recognize the inter-relationship between the various markets in which generation unit owners can participate. In the spot market, unit owners have the option to sell energy and any ancillary services the unit is capable of providing. Unit owners also have the option to sell energy or ancillary services in forward markets. These forward market commitments influence how generation unit owners choose to operate their units and bid them in the spot markets for energy and ancillary services. Finally, unit owners have the option to sell the output of their units to customers located throughout the transmission network. The local market power mitigation mechanism must recognize how a sale in one market impacts the firm's behavior in other markets. Otherwise, the mechanism can lead to market outcomes that yield an unreliable supply of electricity and unjust and unreasonable prices for end use customers.

Third, to maximize the benefits that end use customers receive from wholesale competition, the local market power mitigation mechanism should not allow generation unit owners to use this mechanism to distort market outcomes in their favor. Market prices should be set through competitive bids by generation unit owners. When there are an insufficient number of potential suppliers to a given location in the network, as is the case with local market power, regulatory intervention should occur to protect end users from unjust and unreasonable prices. Setting market prices using a mixture of competitive bids and bids set by regulatory mechanisms can allow generation unit owners to leverage the local market power possessed by some of their units to all of the units they own. Specifically, if bids determined through a regulatory cost-of-service filing must be accepted by the RTO/ITP because of local market power, and these bids are able to influence the market prices received by all units, firms have a strong incentive to alter how they bid and schedule their remaining units to increase the likelihood that these high cost-of-service bids will set market prices. Consequently, mixing regulated and unregulated bids is more likely to subject end use customers to unjust and unreasonable market prices than setting prices using bids submitted by firms facing sufficient local competition.

4. A Mechanism for Local Market Power Mitigation under LMP

This section describes the details of a local market power mechanism for the Commission's SMD wholesale market with locational marginal pricing for managing transmission network congestion that possesses the three properties described in the previous section. This mechanism has two basic components. The first is concerned with local market power mitigation in the spot market. The second mitigates local market power in the forward congestion management market. Each component is first described generally and then illustrated through examples using the 3-node network model given in Figure 1.

The first step in any procedure for local market power mitigation in the spot market is determining whether a firm possesses sufficient market power to require mitigation. I would expect considerable disagreement among stakeholders and regulators over what constitutes local market power sufficient to require mitigation. For this reason, my mechanism specifies what I believe to be is the most generous definition (to generation unit owners) of the level of market power worthy of mitigation. Any firm that is a monopolist for a quantity of energy in a local area must be mitigated for at least the pivotal quantity of energy. There are a number of ways to make this methodology for determining whether a firm possesses significant market power more stringent. My local market power mitigation procedure could also be implemented with a number of these more stringent definitions.

Turning to the example given in Figure 1, the generator at node 3 is a local monopolist with a pivotal quantity of at least 837 MW. The more energy the generator at node 2 injects, the greater is the pivotal quantity of energy for the generator at node 3. This example illustrates a general result in N-node networks--the pivotal quantity of the local monopolist and whether a firm is a local monopolist depend on the configuration of the transmission network, the available capacity of each transmission link and the amount of generation available at each location in the network.

This result suggests the following general procedure for determining whether a firm is a local monopolist during a given hour or day and its pivotal quantity of energy for each hour. At the close of the day-ahead market, after all generation unit owners submit bids showing their willingness to supply during each hour of the following day, the RTO/ITP would first estimate the state of the transmission network for the following day and the level of demand at each location in the network. The RTO/ITP would then determine the total amount of capacity bid into the day-ahead market *at any price* by each generation unit in the control area and all importers into and exporters out of the control area. To determine the extent to which each firm owning generation in the RTO's/ITP's control area is a local monopolist, the RTO/ITP would solve for the minimum amount of generating capacity on a system-wide basis that must be committed for all physical network constraints to be satisfied, given the RTO's/ITP's best guess of demand at each location in the transmission network. Using the capacity commitments resulting from solving this optimization problem, the RTO could determine whether each firm was a local monopolist for any quantity of energy, given the capacity commitments from generation units owned by all other market participants. The minimum quantity of energy over which a firm was a local monopolist would then become that firm's pivotal quantity.

Applying this procedure to the 3-node example given in Figure 1, this procedure would yield a pivotal quantity of energy for generation at node 3 equal to 837 MW. This minimum amount of energy over which generation at node 3 is a local monopolist occurs when generation at node 1 supplies 1,800 and generation at node 2 supplies 0.

Once the pivotal quantities of energy for each market participant have been determined, the RTO would then mitigate the bids submitted by each of these firms for that firm's pivotal quantity of energy. These mitigated bids would not be allowed to enter the LMP process. Instead, the LMP process would assume that each firm was at least supplying the pivotal quantity of energy from its units the RTO/ITP had designated as having any pivotal quantity of energy. By excluding from the LMP process bids from generation units for only their pivotal quantity, this local market power mitigation mechanism prevents cost-of-service regulated bids from being combined with market-based bids in the LMP mechanism, thereby limiting the ability of generation unit owners with local market power to leverage this market power to all of the units they own.

It is possible to use more stringent versions of this mechanism to determine whether a firm possesses sufficient local market power to be worthy of mitigation. For example, the RTO/ITP could require mitigation if two firms can jointly act a local duopolist, rather a single firm acting as a local monopolist. In this case, mitigation would occur if any two firms together owned sufficient capacity to be pivotal. Taking the example in Figure 1, if there are two firms located at node 3 and each of them owns 850 MW of capacity, neither of them would be worthy of mitigation under the local monopolist criterion. However, under the local duopolist criterion, they are jointly pivotal for 837 and therefore worthy of mitigation for a joint supply of at least this quantity of energy. One difficulty with implementing this local duopolist criterion is the procedure used to allocate the pivotal quantity between the two duopolists. In the example from Figure 1, there are a number of ways that the two firms can share the requirement to jointly supply 837 MW of mitigated capacity from their units. For the remainder of my testimony, I will focus of the case of local monopolist mitigation criteria, recognizing that it possible to implement this local market power mitigation methodology with other criteria for determining whether a firm's generation units possess significant local market power.

There are two options for paying the generator for this pivotal quantity of energy. The first option would pay the unit's variable costs, as filed with the RTO/ITP and approved by the Commission, associated with providing the pivotal quantity of energy. Because the firm is eligible to set and earn a market price that can exceed its marginal cost for all energy supplied beyond this pivotal quantity, there is little reason to include an adder to the firm's variable cost for providing this pivotal quantity of energy during a given hour. If on an annual basis, the firm is unable to recover sufficient revenues in excess of the variable costs of production from selling unmitigated capacity into the ITP's/RTO's energy and ancillary services markets, then it should have the option to make a cost-of-service filing with the Commission to obtain additional revenues. Returning to the example in Figure 1, under this scheme the generation unit owner at node 3 would be paid its variable cost for providing 837 MW. It would also be eligible to set and earn the market-clearing price at node 3 for any energy it supplies beyond 837 MW, the amount of energy it must supply in order for system balance and the loop flow constraints to be satisfied.

A second option would be to pay the market-clearing locational marginal price at that unit's node for all of the output it produces. The firm would still be prohibited from setting the market-clearing price with bids from the quantity of energy it must supply to satisfy the system balance and loop flow constraints, but it can set this locational marginal price with bids for any capacity supplied beyond the local energy requirement. For the example in Figure 1, this scheme would pay generation unit owners at node 3 the locational marginal price that excludes its bids for the first 837 MW of its capacity for all energy supplied from its units.

I will first show that under the assumption that all firms exercise no market power (clearly an extremely unrealistic assumption about actual firm behavior), this mechanism yields exactly the same market clearing prices and quantities supplied at each node in the network as the standard LMP mechanism recommended in the SMD. Next, I will discuss the beneficial properties of this mechanism under the more realistic assumption that the firms possess and exercise local market power. When firms possess significant local market power and have strong incentives to exercise it, my mechanism significantly limits their ability to do so and the harm to end users that could result under the standard LMP mechanism.

Consider the network given in Figure 1. This figure contains plots of the marginal cost curve for generation at each node of the network. Node 1 has uniformly lower marginal costs of generation than node 3 and uniformly higher marginal costs of generation than node 2. Let $C_i(q_i)$ denote the total cost of generating q_i at node i and $BP_i(q_i)$ the supply bid price as a function of the quantity demanded at node i. The LMP process computes the prices at each location in the network by minimizing the sum of the areas under the $BP_i(q_i)$ curves over the three nodes, subject to the constraints that the total amount of energy produced at the three nodes equals the demand at node 3 and the transmission capacity constraint on power flows between node 2 and 3 is not violated.

To show that my mechanism does not distort locational marginal prices if all firms display competitive behavior, I assume that no generation unit owner exercises market power, even ones with local market power. This implies that all generation unit owners at all nodes bid the marginal cost curve of each unit they own as that unit's bid supply curve. This bidding behavior implies that the area under the $BP_i(q_i)$ curve equals $C_i(q_i)$, the minimum cost of producing q_i at node i, for all values of q and all nodes. Solving this problem for the marginal cost curves given in Figure 1, assuming no constraints on available generating capacity at any nodes yields: $q_1 = 1{,}643$, $q_2 = 78$, and $q_3 = 915$ and $p_1 = 3.34$, $p_2 = 1.04$, and $p_3 = 5.92$, where p_i is the locational marginal price at node i.

Applying my proposed local market power mitigation mechanism to this example yields the following result. As discussed above, 837 MW is the pivotal quantity at node 3, so that the LMP process assumes that at least 837 MW is supplied at this node. Assuming all firms behave as price-takers for all capacity not subject to local market power mitigation, the LMP mechanism now minimizes the objective function, $C_1(q_1) + C_2(q_2) + [C_3(q_3) - C_3(837)]$, subject to the two constraints given above and the constraint that q_3 is greater than or equal to 837. As discussed above, the last inequality constraint is redundant given the other two constraints. Note that subtracting the value of total costs at node 3 at $q_3 = 837$ MW, $C_3(837)$, is the same as adding a constant to the objective function of the LMP problem described above. The solution to this LMP problem yields the same values for production at each node and the same prices at each node as the LMP problem that assumes all generation unit owners behave as price takers, even the generation unit owner at node 3 that is a monopolist for at least 837 MW at node 3. Consequently, this local market power mitigation procedure replicates the perfectly competitive market outcome when all generation unit owners behave as price-takers for any capacity supplied beyond their pivotal quantity. This result follows from the fact that the constraint that $q_3 \geq 837$ is a direct implication of the load balance constraint, $q_1 + q_2 + q_3 = 2637$, and the transmission capacity constraint, $1/3\ q_1 + 2/3\ q_2 \leq 600$.

Mitigating the bids of generation unit owners for their pivotal quantity of energy and not allowing these mitigated bids to enter the LMP process increases the incentives for generation unit owners to bid as close as possible to their minimum variable cost for any remaining energy they supply to the market, which benefits both system reliability and market efficiency. To appreciate the full implications of these statements, consider the following simplified model of optimal bidding behavior given in Figure 2.

Let DR(p) denote residual demand curve faced by the firm. For each price, p, this function gives the amount of market demand for that hour left to be served by this firm, given the bid supply curves submitted by all other market participants. Because this function depends on the supply bids submitted by other market participants, it can only be known after the prices are set, assuming the bid data is released to all market participants after the fact. Because the firm does not know the realization of the residual demand curve when it bids, the firm instead bids against a distribution of possible residual demand curves. Its bid supply curve, S(p), is set to maximize the expected value of profits, given the distribution of residual demand curves that it faces. Figure 2 considers the simple case of two possible residual demand curve realizations and shows how a firm would construct its expected profit-maximizing aggregate bid supply curve, S(p), under these circumstances.

The simple intuition behind expected profit-maximizing bidding is that the firm would like to submit a bid supply curve that sets the ex post monopoly price for all possible residual demand realizations. In terms of the example in Figure 2, this means that for each realization of the residual demand curve, the firm would like its bid supply curve to intersect the residual demand curve realization at the monopoly price/quantity pair for that curve. A profit-maximizing monopolist facing a known demand curve produces at the level of output where marginal revenue equals its marginal cost. For DR_1 this implies producing at a level of output where the MC curve intersects MR_1. For DR_2 this implies producing at a level of output where the MC curve intersects MR_2. The expected profit-maximizing bid curve therefore passes through the monopoly price/quantity pairs for these two residual demand realizations--(P_1,Q_1) and (P_2,Q_2). Because both residual demand curves have a negative slope, the monopoly price for each residual demand curve exceeds the marginal cost at that level of output. Moreover, the difference between monopoly price and the marginal cost increases with the amount of output supplied by the firm.

The firm's expected profit maximizing bidding strategy can take a very different form if it believes that some residual demand realizations are perfectly inelastic for a positive quantity of energy. This is another way of saying that the firm believes that for some residual demand, it has a positive pivotal quantity of energy. Figure 3 illustrates this case for two residual demand realizations. This figure is similar to Figure 2, but in this case, the low residual demand realization, DR_2(p), is replaced by a high residual demand realization, $DR_3(p)$. For the residual demand curve realization $DR_3(p)$, the firm faces a perfectly inelastic demand at the output level Q_3. Any bid price below the Safety-Net Bid Cap on the spot market will be accepted for at least this quantity of energy. Therefore, for this residual demand realization, the profit-maximizing price quantity pair is (P_3,Q_3), where P_3 is also the Safety-Net Bid Cap on the spot market. If there were no Safety-Net Bid Cap on the spot market, there would be no limit to the profit-maximizing price for this residual demand realization. For the other residual demand curve realization, $DR_1(p)$, the profit-maximizing price/quantity pair is still (P_1,Q_1). However, note that this point is located below and to the right of the pair (P_3,Q_3). Consequently, there is no flat or upward sloping bid supply function S(p) that could connect these two points.

Because all wholesale electricity markets require that firms submit flat or increasing bid supply functions, the firm must decide how to trade off profits between these two residual demand realizations. If the firm perceives the probability of the residual demand realization $DR_3(p)$ as sufficiently high, it may find it will maximize its profits by submitting a bid curve that passes through the point (P_3,Q_3). The decision of firms to exploit this extreme form of market power is one explanation for price spikes in wholesale electricity markets.

If the RTO/ITP implemented my recommended local market power mitigation measure, the firm would be unable to exploit this extreme form of local market power. Instead, the firm would find it expected profit-maximizing to set the price/quantity pair (P_3^M, Q_3^M) given in Figure 3 for $DR_3(p)$. Because (P_3^M, Q_3^M) is located to the right and above (P_2,Q_2) of Figure 2, the expected profit-maximizing bid curve for the firm under this local market power mitigation mechanism would be the bid supply curve $S^M(p)$ given in Figure 3. This example clearly illustrates the incentives for more aggressive bidding by firms with local market power under my proposed local market power mitigation mechanism. Clearly, the bid supply curve $S^M(p)$ is closer to the firm's marginal cost curve than a bid supply curve that passes through the points $(0, P_3)$ and (P_3,Q_3), which is the expected profit-maximizing bid supply curve in the absence of my local market power mitigation mechanism if the probability of residual demand realization $DR_3(p)$ is sufficiently high. Under my suggested local market power mitigation mechanism, firms are only able to exploit local market power that results from facing a nonzero slope to their residual demand curve, rather than the extreme local market power that results when firms face a residual demand quantity that does not change with increases in the market price.

Summarizing the above discussion, there are three important features of my procedure for mitigating local market power in the spot electricity market. First is the mechanism for determining whether a firm is a local monopolist and computing its pivotal quantity if it is a local monopolist. Second is only subjecting to mitigation bids for this pivotal quantity of energy. Third is that these mitigated bids are excluded from the LMP price-setting process.

5. Comparison to PJM Local Market Power Mitigation Mechanism

The Commission has used the PJM market design as a template for much of its SMD. The PJM local market power mitigation mechanism differs from my proposed mechanism in three ways. First, the PJM ISO uses a different procedure for determining whether a generation unit owner possesses significant enough local market power to have its bid mitigated. Second, if it does find that a generation unit possesses local market power, it subjects the price bids for the entire capacity of the unit to mitigation. Third, it allows these mitigated bids to enter the LMP process.

Under the PJM market rules, when a generating unit is determined to possess significant local market power, the ISO automatically mitigates the bids of the unit to one of the following levels: (1) the variable costs of production for the unit plus a ten percent adder, (2) an average of the accepted bids from that unit when it was known not to possess local market power, or (3) a level mutually agreed upon by the market participant and the ISO. In practice, the mechanism that replaces the firm's bid with the unit's variable cost plus a 10% adder is used the vast majority of the time. This mitigated bid is applied to the entire capacity of the unit and the LMP algorithm is then run, with all mitigated bids in place of the actual bids submitted. All units are paid the resulting price at their location for all of the energy they supply to the day-ahead market.

The PJM ISO determines whether a unit possesses sufficient local market power to require mitigation by first looking at the bids across the three major interchanges in the PJM control area. These three interchanges effectively divide the PJM control area into three sub-regions. Bids used to manage congestion across these three interchanges cannot be mitigated. However, if a bid from a generating unit is taken out of merit order on one side of these three interchanges then that bid is mitigated. A generating unit is out of merit order in one of the geographic regions defined by the three interchanges if it is needed to meet a local demand for energy, even though there are lower-priced bids from units in this geographic area that cannot be used because of transmission constraints. These local transmission constraints therefore endow this unit with significant local market power, because it must produce some energy regardless of its bid price. An equivalent way to refer to a unit with significant local market power is as a "must-run unit," because regardless of its bid price, the unit must produce some amount of energy to maintain grid reliability.

It is important to emphasize that a unit owner cannot be out of merit order in a spot market with LMP. In the limiting case of one generating unit at each node in the transmission network, being out of merit order is logically impossible. If more energy is needed at one node in the transmission network, there is only one generating unit that can supply that energy. The supplier bidding the lowest price at that location is also the only supplier at that location, so its bid must be accepted. However, absent mitigation or an enforceable cap on bids there is also no limit to the price that this single supplier can bid for this amount of energy. In the example in Figure 1, the generator located at node 3 can bid any price it would like for its capacity and at least 837 MW would have to be accepted by the LMP pricing process. Consequently, a unit owner can only be out of merit order in a geographic area that contains multiple nodes. It has this status because transmission constraints within this region prevent the ISO from accepting the lowest priced bid in a region containing multiple nodes. These local transmission constraints prevent energy injected at the node from the lowest-priced available generating unit from serving the demand increase in this geographic area.

Returning to the example in Figure 1, suppose the 3-node network in this example was part of a larger network and these three nodes were considered the relevant geographic area over which the out-of-merit-order determination would be made. Suppose the units at node 3 submitted any bid supply function above their variable operating costs. Under the PJM mechanism, the generator located at node 3 would be considered out of merit order because at least 837 MW must be accepted from these units, rather than from the low-priced units at node 2. Depending on the amount of generating capacity at each node, the level of demand at node 3, and the bids that units at node 1 submit, the units at node 1 could also be out of merit order, and therefore subject to mitigation under the PJM mechanism.

Because the PJM LMP procedure includes mitigated price bids of the entire capacity of the unit in the LMP process, this creates incentives for firms owning portfolios of generating units to bid or schedule these units to increase the likelihood that units with local market power will set higher prices for all units. This logic is useful for understanding how firms are able to raise prices to very high levels during tight system conditions in the PJM market. As discussed for our 3-node example in Figure 1, under tight system conditions, i.e., when available generation capacity exceeds system demand by a small amount, the number of units with bids that will be mitigated is likely to be much larger. Moreover, these additional mitigated units are likely to have higher cost-based bid levels because they tend to be mid-merit or peaking units with higher heat rates. Consequently, the PJM procedure of including bids from generators with local market power (even at mitigated levels) in the price-setting process allows the remaining units selling into the market to bid significantly higher because the firms that own these units know that higher mitigated bids must be accepted.

The logic of the above example continues to hold if generating unit owners only know the probability distribution of the quantity of energy that will be accepted from the units with local market power. This still enables the other units to raise their bids, because they know with a very high probability that during tight system conditions some quantity of energy from these high-cost units must be accepted and paid prices greater than or equal to these mitigated bid levels.

It is also possible to construct examples where the local market power mechanism PJM uses artificially depresses prices at certain locations in the transmission network because the entire capacity of the mitigated unit enters into LMP process with a mitigated bid, rather than just the pivotal quantity, as is the case with my local market power mitigation mechanism. All of these upward and downward price distortions occur because the PJM local market power mitigation mechanism allows cost-of-service regulated bids for local reliability energy and competitive bids to determine jointly all locational marginal prices. Therefore, to ensure that consumers fully realize the benefits of wholesale competition, the local market power mitigation measure adopted as part of the Commission's SMD should not allow mitigated bids to enter LMP process.

6. Flexibility in Design of Proposed Local Market Power Mitigation Mechanism

Although my proposed mechanism does not allow the mitigated bid to enter the LMP process, as discussed above, it is flexible in terms of how this firm is paid for the pivotal quantity of energy it provides. Specifically, the firm could be offered the option to receive the resulting locational marginal price at its location for the pivotal quantity energy, instead of its variable costs. This firm could also be offered its variable costs plus a 10 percent adder or any other adder that the Commission might deem reasonable. However, because the firm is allowed to bid the remainder of the capacity of its units beyond the pivotal quantity and these bids will enter the LMP process, the logic for allowing firms to earn an adder for supplying the pivotal quantity of energy is less compelling. Consequently, my recommended alternative is to allow the firm to elect on a daily basis whether it will receive its variable costs or the resulting LMP times the pivotal quantity of energy for providing what is effectively a regulated service.

My mechanism also allows a unit that is determined to possess local market power in the day-ahead market to participate as a price-setting bidder in the real-time market. However, if it is determined in real-time that the unit possesses local market power, it is essential that this unit not be able to set prices with its pivotal quantity of energy. This should be the case because there is no limit to what that unit can bid and still be accepted to supply energy beyond this pivotal quantity. As an example of how a firm might be allowed to set prices in the real-time market, suppose that by scheduling 100 MWh in the day-ahead market in a local area, a firm supplies its pivotal quantity of energy requirement from the day-ahead market. In other words, the RTO/ITP has determined that this firm is no longer a local monopolist if it supplies 100 MWh or more of energy. Under these circumstances, it would not be harmful to competition for the RTO/ITP to allow this firm to submit market-based bids to supply beyond 100 MWh of energy from these units in real-time. This is because the firm faces effective competition from importers into this local area for energy beyond the 100 MWh it is committed to supply in the day-ahead market.

By treating units with local market power as price-takers in the LMP-setting process, the incentive the firms have to alter their bidding and scheduling behavior to take advantage of the fact that these cost-based bids will be accepted with certainty would be eliminated. Under certain system conditions, these cost-based bids can become floors for the bids that all other firms will submit and can therefore significantly inflate or depress locational prices. In the 3-node example in Figure 1, if the firms at nodes 1 and 2 knew that the bid entering into the LMP process for 900 MW of capacity at node 3 would be mitigated to its variable cost plus a 10 percent adder, the unit owners at nodes 1 and 2 would have less incentive to bid close to their marginal cost curve for supplying energy at their respective nodes. Treating the pivotal quantity of output from the firm at node 3 as a regulated local reliability service and not allowing the bids for this pivotal quantity (even at mitigated levels) to influence the LMP-setting process limits the ability of firms to exercise significant local and portfolio-wide market power.

Treating the output of these units as a regulated service is consistent with the fact that there is not sufficient competition for the provision of energy at that location in the network. Suppose that in the 3-node example of Figure 1, the total capacity at node 1 was equal to 1,200 MW. If a single firm owned this capacity, then it would be a local monopolist for 837 MW. If two equal sized firms owned this capacity, then each firm would be a local monopolist for at least 237 MW (the difference between 837 MW and capacity owned by the other firm, in this case 600 MW). If we assume three equal-sized firms own this capacity, then each is pivotal for 37 MW (the difference between 837 MW and amount capacity owned by the other two firms, in this case 800 MW). Finally, if there are four equal sized firms, then none of them is pivotal. In this case, we can most likely rely on competition among firms at node 3 to set competitive prices at this location. However, given the network configuration and cost structure in Figure 1, even competitive bidding by all firms will lead to significantly higher locational marginal prices at node 3, than the other two nodes. As discussed in Section 1, this market outcome is purely a function of implementing a wholesale market using LMP on a network configuration designed to be optimal for the vertically integrated utility. One goal of the second part of my proposed local market power mitigation mechanism is to address this transition period issue between the vertically integrated utility and the wholesale market regimes.

7. Using CRRs to Protect Consumers from Local Market Power

As discussed in Section 3, commitments made by generators in forward markets can have a dramatic impact on their behavior in subsequent markets. Consequently, the second part of my local market power mitigation mechanism focuses on the issue of how to allocate Congestion Revenue Rights (CRRs) among market participants to ensure that they are used as intended, rather than used to enhance the ability of generation unit owners to exercise local market power. This discussion does not address deviations from the PJM simultaneous feasibility approach and grandfathered contracts. These topics are beyond the scope of this evaluation and must be dealt with as a transitional issue best determined by each RTO/ITP.

As has been studied by a number of authors, the method of allocating CRRs among market participants can enhance or reduce the ability of firms owning generating units to exercise market power in subsequent markets.[4] This result has implications for how CRRs are allocated among market participants. Suppose an auction mechanism is used to allocate CRRs among generators, LSEs, and other market participants. One would expect that the market participant that derives the greatest economic value from owning a given CRR would purchase it. For most CRRs, firms owning generation units, because they have the ability to schedule and bid their units, would derive the greatest economic value from them.

CRRs are likely to be far more valuable to generating unit owners that own large portfolios of generating units than to LSEs for the following reasons. Firms with large portfolios of generating units possess the flexibility to bid, schedule and operate their units to impact the level and location of congestion in the transmission network. In contrast, because of the current insensitivity of virtually all end-use demand to the real-time price of electricity, large LSEs have little, if any, true flexibility to bid and schedule their loads to impact the level and location of congestion in the transmission network. As a consequence, LSEs value CRRs only because they provide insurance against congestion charges they have little ability to impact. Portfolio generation unit owners also value CRRs for this reason. But they also value them for their ability to increase the returns to exercising local market power by how they schedule their entire portfolio of generation units. Firms owning portfolios of generation units can increase the frequency and magnitude of the congestion revenues they receive from owning CRRs, depending on how they operate these units. In addition to valuing CRRs for their ability to insure against locational price differences, generation unit owners can earn additional profits from CRRs because one of the mechanisms these firms use to exercise local market power is to impact congestion charges through how they bid and schedule their units. Consequently, the value of CRRs (in terms of the stream of future congestion revenues) to LSEs is likely to be significantly less than the value of these same CRRs to large portfolio generation owners, because CRRs enhance the ability of these firms to exercise local market power. This same logic implies that CRRs are potentially more valuable to

[4] Joskow, P. and Tirole, J. (2000) "Transmission Rights and Market Power on Electric Power Networks, *Rand Journal of Economics*, Volume 31, Number 3, pp. 450-487, provide an comprehensive discussion of these issues.

LSEs that own some generating units in their service areas relative to LSEs that do not own any generating units in their service area.

Any SMD should be structured so that CRRs benefit, to the greatest extent possible, LSEs that use them as a hedge against locational price differences between the LSE's source of power and the location where this electricity is withdrawn from the transmission network. Generation unit owners should not use CRRs as a tool to exercise unilateral market power by artificially increasing the frequency and magnitude of congestion in the RTO's/ITP's network. A CRR allocation mechanism that does not account for local generation ownership or long-term commitments to energy from local generating units by LSEs makes this socially inefficient use of CRRs more likely to occur.

This logic yields two very important principles for market design. First, CRRs should be allocated to LSEs. By allocating CRRs to LSEs, an RTO/ITP limits the opportunities and incentives generation unit owners have to exercise local market power by using their units to cause congestion that artificially inflates the revenues they receive from their CRR holdings. This mechanism increases the incentives generation unit owners have to submit unit-level bid supply functions close to each unit's marginal cost function. Second, the CRR allocation process should account for the fact that a generation-unit-owning LSE or one with long-term commitments from local generation units needs fewer CRRs to achieve the same level of protection from congestion charges. These long-term entitlements to the output of local generation units provide a physical hedge against congestion charges, so that the LSE needs fewer CRRs than an equivalent LSE having none of these long-term rights to the output of local generation units.

How CRRs should be allocated between LSEs with local generation and LSEs with no local generation depends on how one interprets the statement "providing the maximum protection to consumers from congestion charges." I discuss this issue in more detail below, but clearly this allocation mechanism should reward the LSE for owning this local generation or the long-term right to local generation. Conversely, the generation-owning LSE should not be allocated so many CRRs that it would be a profit-maximizing strategy for the firm to use these CRRs as a profit center by scheduling the portfolio of generation units the firm owns to increase the magnitude and frequency of congestion charges it receives from its CRRs. It is important to emphasize that during many hours of the year there are congestion charges that cannot be hedged because insufficient CRRs have been allocated to a given transmission path relative to power flows on the path during that hour. For this reason, allocating too many CRRs to an LSE that owns local generation has the potential to impose significant harm on the LSEs that do not own local generation. Consequently, the CRR allocation process must balance the need to protect the end users of the LSE that owns no local generation against the need to reward the LSE with local generation for owning and operating these units.

Allocating CRRs to LSEs maximizes the likelihood that no end user experiences substantial financial harm as a result of the transition to wholesale competition. Even though customers located in areas with limited transmission capacity and expensive local generation will see high locational marginal prices, because they are allocated CRRs, they only face a residual level of congestion risk due to the uncertainty in the hourly transmission capacity availability relative to the amount of CRRs that the RTO/ITP has allocated to that transmission path. However, it is important to note that depending on the process used by the RTO/ITP to allocate CRRs among the LSEs, some them could still face a non-trivial level of residual congestion charge risk. The CRR allocation process used the by RTO/ITP should offer sufficient year-to-year flexibility to address this potential source of end use customer harm.

Even though CRRs are allocated to LSEs, if load growth occurs, the incremental load will face the prospect of a high locational marginal price if no new generation entry or transmission capacity expansion takes place. These prices are needed to provide the appropriate economic incentives for both of these supply responses to occur. These prices also provide strong economic incentives for the development of demand-response response programs in this local area.

In summary, by allocating all CRR capacity in the manner recommended above, the Commission has a better likelihood of realizing its goal of seeing that all U.S. end users benefit from wholesale electricity competition. Existing loads will, in the aggregate, be hedged against congestion charges to the maximum extent possible, while incremental load, as well as those who may build new generation and transmission facilities, will see the appropriate locational marginal prices. However, I would like to emphasize that it is impossible to hedge all possible transmission congestion, so the RTO/ITP must set the level and geographic configuration of CRR capacity and operate the transmission network to prevent undue harm to end users from quantity of congestion charges that remains.

Arguments that CRRs should be allocated to those who "value them most" fail to recognize that a major reason why one firm may value CRRs more than another firm is because it can create significantly more congestion revenues from the same quantity of

CRRs by its scheduling and bidding behavior. Ignoring this reality in the design of a CRR allocation process can result in market outcomes with significantly higher levels of congestion in the transmission network and thereby increase the likelihood that end use customers will be subject to unjust and unreasonable wholesale prices.

Depending on how many CRRs a generation-unit-owning LSE has, it could face fundamentally different incentives for scheduling and operating units in its service area. For example, if an LSE owns or controls both substantial local generation and a substantial amount of CRRs covering transmission paths into its service area, it could maximize its profits by scheduling these units to increase, rather than reduce, congestion. If the motivation for awarding CRRs to LSEs is to provide a hedge for the congestion-cost risk associated with meeting their load obligations, but not to enhance the ability of portfolio generation unit owners to exercise local market power, then an LSE that owns generating capacity or has long-term commitment to the output from units located near its loads should be treated different from an LSE that owns no local generating capacity.

For example, for the same CRR holdings, an LSE with a peak load of 1,000 MW and 400 MW of generation in its service territory (or an equivalent long-term contract for energy from these units) is significantly less exposed to the congestion charge risk than a LSE with the same peak load but no local generation capacity or long-term contract for energy from these units. The LSE that owns generation already possesses a physical hedge against congestion costs in the form of its local generation. The LSE that owns local generation and significant CRRs into its service area may instead decide to use these local generating units to maximize the profits that it earns from its CRR payments at the expense of slightly lower generation profits. For example, an LSE might be willing to exercise local market power by allowing more frequent and higher congestion charges on an interface covered by CRRs that the firm owns, even though it may be more efficient for the market as a whole for the firm to operate its local generating units to limit these congestion charges.

Netting out some fraction of the local generation that an LSE owns or controls through long-term contractual commitments limits the incentive a generation-owning LSE would have to use its CRRs to exercise local market power by increasing the frequency and magnitude of congestion charges to its own financial benefit, but to the detriment of the market. The RTO/ITP should set this fraction of the LSE's local generation commitments netted against its load obligations to balance a number of competing goals. The owners of these local generating units should be given strong incentives to operate these units when it would be economic to do so (but not to operate them when it would uneconomic to do so), and strong disincentives to use their ability to control these units to increase congestion. Similarly, LSEs that own no local generation or long-term contracts for local generation should not be subsidized as a result of the operation of these units. One possible implication of this scheme is that more efficient (lower-cost) local generation units would have a larger fraction of their capacity netted out against the LSE's local load obligations.

The details of this allocation mechanism should differ across RTOs/ITPs, depending on a number of factors such as the amount and location of hydroelectric power in the RTO/ITP control area(s) and the dependence of the LSEs in the RTO’s/ITP’s region on imports to meet their load obligations. However, the basic recommendation that CRRs are allocated to LSEs, with fewer CRRs allocated to LSEs that own or have verifiable long-term contracts for the output from local generation units, should be followed. If we define the maximum benefits to end users from allocating CRRs to LSEs as minimizing total wholesale energy costs, then there are a number of more concrete mechanisms that can be derived for allocating CRRs among LSEs. However, all of these mechanisms imply giving fewer CRRs to LSE owning local generation relative to an LSE serving the same pattern of annual load that owns no local generation.

The example in Figure 1 can be used to illustrate this CRR allocation mechanism. Suppose that there are two equal size LSEs at node 3. LSE1 owns all 1,200 MW of the expensive generating capacity at node 3. LSE2 owns no generation at node 3. Even though both of these LSEs serve 1,318.5 MW of load, it makes very little economic sense to award them the same quantity of CRRs. This CRR allocation would incent LSE1 to operate its local units considerably less intensively than would be least cost on a system-wide basis. For this reason, LSE2 should receive significantly more CRRs than LSE1. But the RTO/ITP should not simply net out the 1,200 MW of local capacity from LSE1's load of 1,318.5 MW and only award it 118.5 MW of CRRs. This would lead to economically inefficient market outcomes, because LSE1 would run its local units far too intensively for their variable cost. Because the 1,200 MW of capacity owned by LSE1 is far more expensive than the capacity at nodes 1 and 2, the RTO/ITP should allocate sufficient CRRs to LSE1 so that it will have a strong incentive to use its local generation when (and only when) congestion charges into node 3 are likely to be the greatest and local energy prices the highest.

Under this mechanism, an LSE that owns local generation has strong incentives to limit congestion charges rather than increase congestion charges because of the increased profits it earns from its CRR holdings. Any CRR allocation mechanisms implemented in as part of the Commission's SMD must align as closely as possible, the profit-maximizing incentives of the generation-unit-owning LSE with the goal of enhancing rather than detracting from efficient operation of the transmission network.

8. Ongoing Consumer Protection Through An Independent Market Monitor

Perhaps the most important lesson learned from the experience of the almost four years of wholesale electricity competition in the U.S. is that, no matter how much planning and analysis goes into a market design, there are always market design or local market power problems that were not anticipated at the start of market. All of the wholesale markets currently operating in the U.S. are now dealing with a number of market design issues. Many of these problems must be solved in a timely manner or significant harm to end users could occur and the Commission would not achieve its goal of ensuring that U.S. end users benefit from wholesale electricity competition.

The RTO/ITP and the Commission should be as diligent and comprehensive in performing the best possible prospective market power analyses before the start of the wholesale market. All possible prospective market power mitigation remedies, including the divestiture of generation units, should be considered and implemented before the start of the market. However, unanticipated market design and market power problems are still likely to arise. These problems must be efficiently identified and analyzed through effective market monitoring.

Several features of wholesale electricity markets make this a very complex task. Regional electricity networks across the U.S. are extremely idiosyncratic. The details of market rules across the currently operating RTOs/ITPs differ in a number of important respects. The extent of these differences increases or decreases depending on market rule changes made at each RTO/ITP. Although the Commission's SMD process will reduce the broad differences in market rules across RTOs/ITPs, there will remain significant differences among RTO/ITP market rules, depending how the Commission's SMD is implemented in that region. For these reasons, effective market monitoring requires both daily interaction with the RTO/ITP management and market participants and a forward-looking, long-term perspective that anticipates market design problems and takes the appropriate action before they become larger, more serious problems.

In some of the ISOs/RTOs that currently exist, much of this daily interaction takes place within the ISO's own Market Monitoring Unit (MMU). However, there is a very important additional role for a market monitor that is often extremely difficult for the ISO's MMU to play. Members of the ISO's MMU are employees of the ISO, and as such, may find it difficult to perform analyses or publicly state conclusions that may reflect negatively on the ISO's management. There is also considerable room for discretion in the interpretation of each ISO's tariff, and the details of how market rules are implemented by the ISO management can significantly impact the costs to end users of any potential market design flaw or result in the exercise of significant local market power. Consequently, it is essential that the market monitoring entity has no financial stake in any market participant and not be an employee of the ISO/RTO.

A market monitor or market monitoring committee that is financially independent of the RTO's/ITP's MMU can be a source of unbiased expert advice and analysis of market design issues. If it is financially independent from the RTO/ITP and reports directly to the Commission, its goals should be aligned with the Commission's goal of making the market work as efficiently as possible so that all consumers benefit from wholesale electricity competition. Moreover, this independent market monitor should not only oversee and report on the operation of the ISO's/RTO's energy and ancillary services markets, it should also provide analyses of the performance of the RTO's/ITP's own operators and management, an extremely important task that the RTO's/ITP's MMU cannot carry out.

To be most effective, the independent market monitor must have access to all data collected by the RTO/ITP on the characteristics of the transmission network, generating facilities located throughout this network and all of the data submitted by firms to the RTO/ISO and produced as part of the operation of the RTO's/ITP's markets. Moreover, the independent market monitoring entity should have the ability to request data from market participants that is necessary to carry out its market monitoring mandate subject to the constraint that this request can be justified on a cost/benefit basis. This means that potential costs to the firm from compiling and providing this information to the RTO/ITP can be justified in terms of the potential benefits to consumers from the independent market monitor from having access to this data.

In my own experience, I have found it very useful for an ISO's MMU to work closely with the independent market monitor. The internal MMU can then take advantage of the expertise on market design or other topics relevant to market monitoring that the independent market monitor might have. Conversely, the independent market monitor can take advantage of the intimate knowledge of the RTO's/ITP's day-to-day operating procedures and market outcomes possessed by MMU's staff.

Finally, this independent market monitor should prepare periodic reports to the Commission on market performance and the performance of the RTO/ITP, with recommendations for correcting any market design flaws, local market power problems, or system operation problems that it identifies. Through this interactive and collaborate process with the RTO's/ITP's MMU, the RTO/ITP management, and the relevant Commission staff, an independent market monitoring committee or monitor can be an integral part of an effective market monitoring process to protect end users from unjust and unreasonable prices on an ongoing basis and to increase the transparency of the RTO/ITP's processes to market participants.

Although it may seem that having both an internal and external market monitoring entities would create unnecessary duplication of effort and expertise, I have found that having these two entities solves two very important problems associated with the market monitoring process. The first is what I call the "two master problem." No matter how well-intentioned a single internal or external market monitor might be, there will be many circumstances when it will be faced with the choice of serving the RTO/ITP management, the Commission, or its own self-interest. If there is a single MMU located in the RTO/ITP, then it may favor the wishes of the management of the RTO/ITP at the expense of the Commission or consumers. Conversely, an independent market monitor that reports directly to the Commission may not serve the interests of the RTO/ITP management.

I have also found that when an outside market monitor has access to an MMU, it can help bridge what I call the "gains from trade within the organization problem." To be most effective, a market monitor should interact with the market operators, the system operators, and the compliance section of the RTO/ITP on a frequent basis. If the market monitor is not an employee of the RTO/ITP, then there will be a tendency on the part of other employees of the RTO/ITP not to interact with the market monitor. If they do so, they might disclose a shortcoming of the ISO operations that causes the Commission or stakeholders to question the ability of the ISO management. Having a set of market monitors that are also employees of the ISO can enhance the opportunities for gains from trade of expertise between them and employees of other parts of the ISO, to the benefit of market participants and end users. The division of labor among system operation, compliance, client relations, legal affairs and market monitoring is often blurred. Cooperation between all divisions of the RTO/ITP is essential to making markets work to benefit consumers. The market monitoring process is a vital component of virtually all aspects of the operation of the RTO/ITP. If there is a single market monitoring entity financially independent of the RTO/ITP, employees in other divisions of the RTO/ITP might be less likely to interact as freely with the market monitoring entity, which will have potentially detrimental impacts on market efficiency. Consequently, I think that if the independent market monitor can have access to an inside market monitoring group that is composed of employees of the RTO/ITP, this would maximize the possibilities for gains from trade between the market monitoring function and other parts of the RTO/ITP.

I have found that the same gains-from-trade logic applies to the case of the independent external market monitor. Having both an internal MMU that is employed by the RTO/ITP and an external independent market monitor also solves the "two-masters problem" because the internal MMU works for the RTO/ITP, and external monitor for the Commission. There are also potential gains from the trading of expertise between the internal and external market monitor. The external monitor can be more outward focused, looking across markets, with a long-term view, whereas the internal market monitor can focus on the day-to-day operation of the RTO/ITP.

____
Frank A. Wolak

SUBSCRIBED AND SWORN TO BEFORE ME, THIS _____ DAY OF _____________, 2002

____
NOTARY PUBLIC

Figure 1: Three-Node Network Model

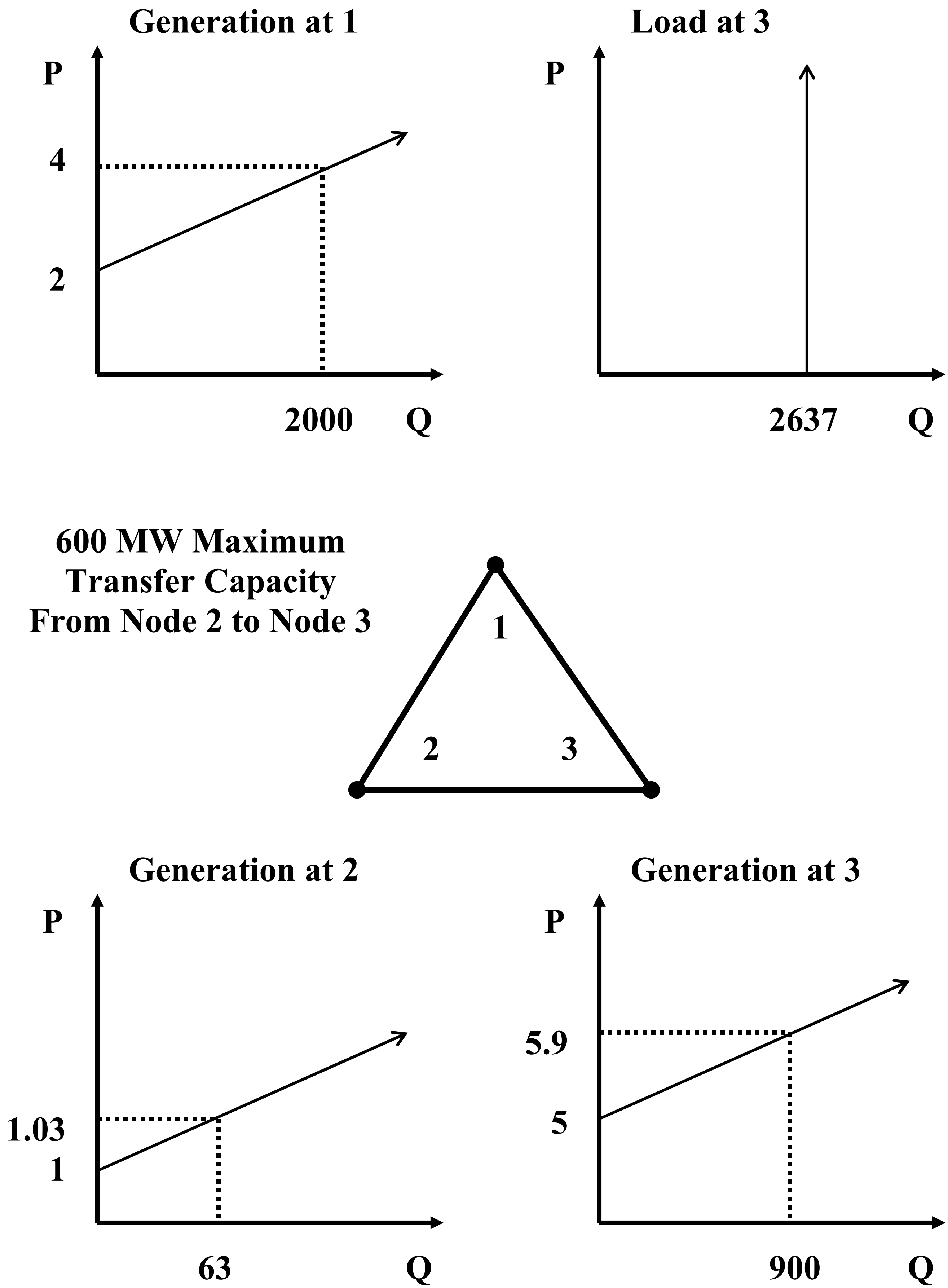

Figure 2: Bid Curve For Two Residual Demand Realizations

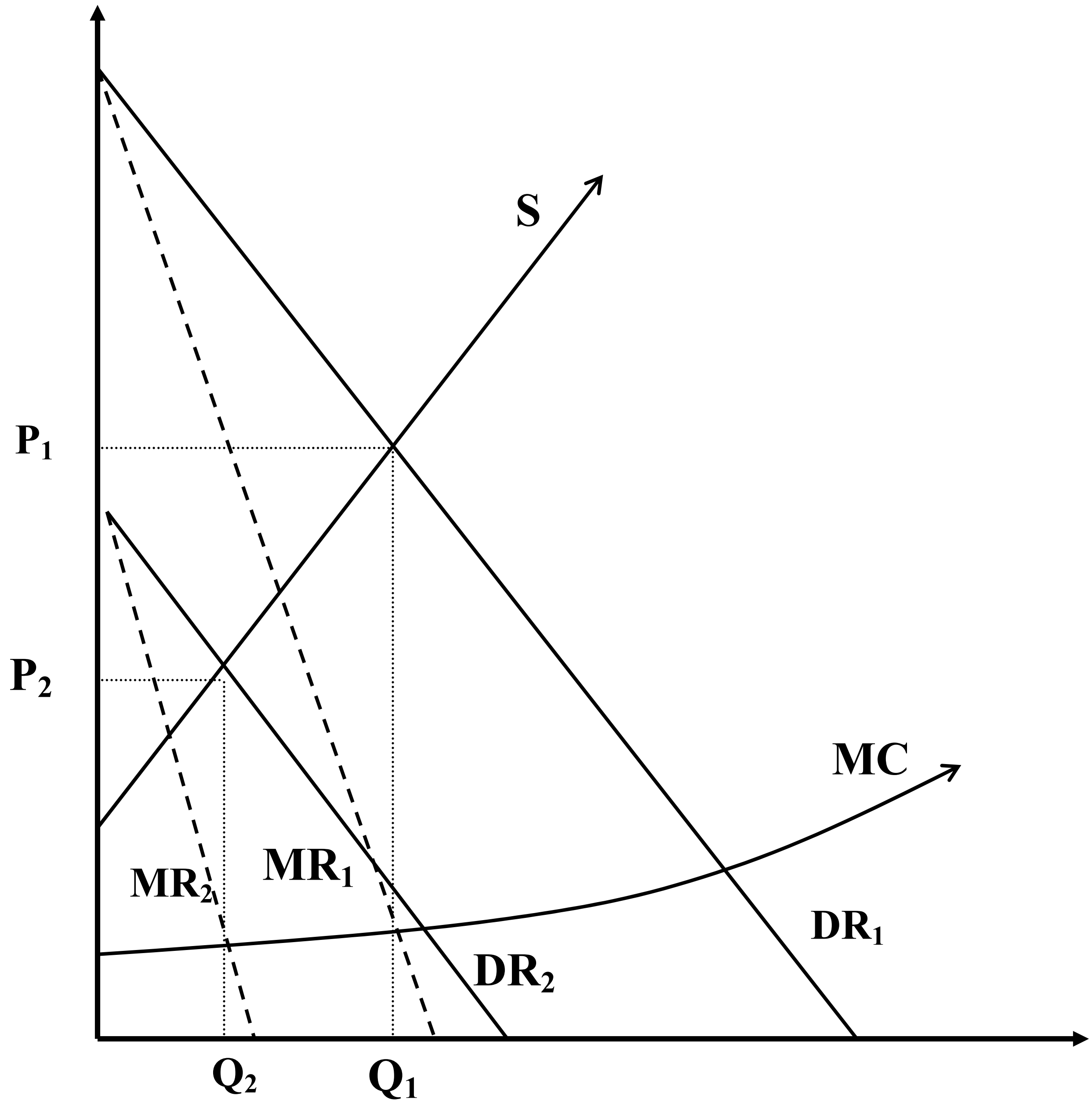

Figure 3: Bidding With One Pivotal Residual Demand Realization