

A Standard Market Design for Regional Transmission Organizations

John D. Chandley¹

It is time for the Federal Energy Regulatory Commission (FERC) to define the principles for a standard electricity market design and to begin consistently applying those principles to the market rules now being developed by Regional Transmission Organizations (RTOs). FERC's recent RTO orders make it abundantly clear that the fundamental purpose of forming independent RTOs is not merely to "operate the grid," nor only to ensure non-discriminatory access to essential grid facilities and services. In addition to these undisputed RTO responsibilities, an essential function of RTOs is to create and operate RTO-coordinated markets.

Competing generators, customers/buyers and traders will use RTO-coordinated markets to support bilateral contracting and a wide range of commercial market transactions. The markets will also serve a broader national objective of pursuing the public interest in the benefits of market competition. RTOs will use these markets to maintain regional reliability through system balancing, congestion management and provision of operating reserves and security services, thus assuring a safe and dependable electricity system and a reliable supply of electricity. The success of RTO markets is therefore critical to system reliability, as well as national economic well being.

Market experiences in various regions have demonstrated beyond debate that poor market designs can expose affected states and participants to enormous risks. Given these risks, FERC can no longer afford to avoid the hard job of defining a workable, robust and efficient market design. And because the electricity grid functions across huge interconnected regions – as one coordinated system -- the market rules that apply across the country must be fundamentally compatible, as well as technically sound. In short, if the nation is to embrace competitive markets as the appropriate structure for the electricity industry, FERC must define and articulate the principles for a standard market design and insist that RTOs use those principles to develop their market rules. This paper describes a set of foundation principles for such a standard design.

FERC's Policy Orders

FERC has been attempting for almost a decade to restructure the electric industry to foster competitive electricity markets. After pursuing this objective under the existing industry structure through Orders 888 and 889, FERC has now concluded that it must create RTOs to achieve the goal of fostering competitive markets. In Order 2000,² its "Millennium Order," FERC called upon jurisdictional transmission-owning utilities to create RTOs and to transfer

¹ The author is a member of the LECG market design group led by Harvard's William Hogan and LECG's Scott Harvey. The paper is based on electricity market design principles developed over several years by Drs. Hogan and Harvey, other members of the design group, and other international experts in market design. The paper was commissioned by Commonwealth Edison, with the support of a number of parties interested in a clear statement of market design principles. The paper does not necessarily represent the views of Commonwealth Edison or any other party. While members of the LECG design group and others contributed helpful comments on the paper, any errors are attributable solely to the author.

² Regional Transmission Organizations, Order 2000, 65 Fed. Reg. 809 (Jan. 6, 2000), FERC Stats. & Regs. 31,089 (1999), [order on reh'g](#), Order No. 2000-A, 65 Fed. Reg. 12,088 (March 8, 2000), FERC Stats. & Regs. 31,092 (2000), [rehearing pending](#). Hereafter cited as Order 2000. All page references in the footnotes are to the Order published on the FERC web site.

control of grid operations to their respective RTOs. The Order also requires each RTO to meet certain minimum requirements to support competitive markets, but more guidance is now necessary.

FERC's Order 2000 defines the minimum requirements that every RTO must meet, while suggesting (without mandating) preferred approaches for meeting each requirement. In particular, Order 2000 requires that every RTO provide at least the following support for regional markets:

- An RTO must provide a real-time balancing market and ensure that all parties have non-discriminatory access to this market. Order 2000 notes that bid-based markets coordinated by an RTO and settled at market-clearing prices provide a proven, workable approach.³
- An RTO must provide market-based mechanisms to manage congestion within the region and to deal effectively with transmission loop flows, rather than rely on administrative (TLR) curtailments. FERC appears to be uncertain and less prescriptive here, but Order 2000 notes that approved, working markets have used the bid-based dispatch that provides system balancing to also provide a market-based redispatch alternative to relieve congestion.⁴
- An RTO must price congestion and imbalances efficiently, so that generators and other parties have appropriate price signals to encourage efficient short-run operations and long-run investments. Here, FERC notes its prior approvals for pricing mechanisms based on locational marginal pricing.⁵
- An RTO must offer tradable transmission rights that allow parties to hedge locational differences in energy prices resulting from congestion. These rights must support efficient regional dispatch and provide efficient incentives. The language implies a need for "financial" rights that function as price hedges rather than prerequisites for grid access that might preclude alternative grid uses or undermine an efficient dispatch.⁶

Last December, in response to Order 2000, utilities and affected parties in over a dozen regions of the United States proposed new RTOs or asked to have existing Independent System Operators (ISOs) approved as conforming RTOs. FERC has now issued initial decisions on these proposals, rejecting some proposals for various reasons and granting conditional approval to others. Importantly, on July 12, 2001, FERC issued a series of RTO decisions that, instead of leaving the number and size of RTOs to be defined through discussions among the affected parties, strongly encourage the formation of four RTOs:

³ Order 2000, at pp. 423-425 and 633.

⁴ Order 2000, at pp. 333-334, 381-382.

⁵ Order 2000, at pp. 332, 382.

⁶ Order 2000, at 333, 489.

We favor the development of one RTO for the Northeast, one RTO for the Midwest, one RTO for the Southeast and one RTO for the West.⁷

FERC's desire to increase the size and limit the number of regional markets reinforces the need for a standard market design. FERC appears to recognize that its market objectives could be frustrated if there were numerous regional markets, each with a different set of parochial market rules that impose barriers to inter-regional trading. Yet the same argument applies whether there are four or 14 regional markets. The simple truth is that robust inter-regional trading and competition are unlikely to flourish if neighboring RTOs use incompatible market designs, disjointed dispatch and scheduling rules and inconsistent pricing schemes.

The broader policy objective cannot simply be a set of "seams agreements" for working around the barriers created by inconsistent regional markets, even if there are only four such regional markets. Rather the policy goal should be to facilitate seamless inter-regional trading to the degree practicable. While even fewer RTOs may be too great a technical and policy challenge for now, multiple RTOs could at least minimize seams issues by implementing a set of common market design principles that allow trading to occur across inter-regional boundaries with a minimum of administrative transaction costs and disruptions. Logically, applying a standard market design is a necessary condition, though not a totally sufficient one, for achieving these goals.

Defining a Best Practices Standard

In its July 12 Orders provisionally approving the PJM RTO proposal and rejecting the New York and New England proposals in order to encourage the formation of a single Northeast RTO, FERC indicates that it expects to see the Northeast region operating under a common set of market rules based on "best practices" from the existing ISOs:

PJM's RTO proposal can serve as a platform for the formation of one RTO in the Northeast . . . Along these lines we encourage the three ISOs to look at the best practices in all three ISOs to develop market rules for a Northeast RTO. While we would expect that PJM will be the platform for forming a single Northeast RTO, we also would expect the RTO proposal to incorporate the best practices of the NYISO and ISO New England.⁸

Similarly, in calling for the formation of a single RTO for the Southeast region, FERC also encourages the region to develop a common set of market rules based on best practices:

The [congestion management] mechanism ultimately proposed by Grid South to replace the interim mechanism should either implement the best practices from among the existing mechanisms currently in use by other grid operators or explain why its proposal is superior to the industry's existing best practices.⁹

⁷ See, e.g., FERC, "Order Provisionally Granting RTO Status" (for the PJM RTO), issued July 12, 2001, in Docket No. RTO1-2-000, page 2. Hereafter, cited as PJM RTO Order.

⁸ PJM RTO Order, page 13.

⁹ FERC, "Order on Compliance Filing and Status Report," issued July 12, 2001 in Docket Nos. RTO1-74-002 and RTO1-74-003, page 23.

FERC has not determined what these “best practices” include. While Order 2000 lists several minimum requirements for an RTO’s market support, it gives flexibility to each RTO in developing market rules that satisfy the Order’s minimum requirements. And on approaches to congestion management, the Order declares FERC to be open to innovation and experimentation. However, there may only be a narrow range of design choices within the “existing mechanisms currently in use.” The most obvious choices include:

- (1) *The current pre-ISO approaches.* In regions that do not yet have independent system operators, trading must rely on utility controlled dispatches, restricted imbalance services, contract-path transmission scheduling and (in the Eastern Interconnection) administrative curtailments (“unscheduling”) using Transmission Line Loading Relief (TLR) procedures. There are no spot markets, no ability to buy through congestion, and no workable systems for tradable transmission rights
- (2) *The California market design.* This uses an ISO to coordinate bid-based spot/balancing and congestion management markets, while using zonal pricing to price congestion and spot market transactions. The ISO offers inter-zonal transmission rights (a form of flowgate rights) that function more or less as financial hedges against inter-zonal congestion charges.
- (3) *The Northeastern market designs.* These markets also use ISOs to coordinate bid-based spot/ balancing and congestion management markets (as well as day-ahead energy and transmission markets), but they use nodal locational marginal pricing to price congestion and spot market transactions. The ISO offers point-to-point financial transmission rights that function as hedges against congestion-based spot prices and point-to-point congestion charges. Parties willing to pay the marginal cost of redispatch can buy through congestion.
- (4) *A “hybrid” market design.* These might use some combination of the Northeastern designs for real-time congestion management using nodal pricing, and flowgate transmission rights to help manage or hedge congestion in forward periods. There are no designs of this type currently in operation, but variations are being considered by some emerging RTOs.

Order 2000 has already rejected the existing contract-path/TLR approach as unequal to the challenges of supporting a competitive market. It also seems unlikely that FERC would readily accept another design premised on the uniform congestion pricing schemes that it has already found to be problematic in PJM (1997) and New England (1999) or the inter-zonal/intra-zonal distinctions that it found to be fundamentally flawed in California (2000). There are no operating markets using any of the proposed “hybrid” market designs at this time, though several are being considered. While this would appear to substantially narrow the choices, FERC has been careful not to mandate any particular approach:

... we reiterate that, while LMP is an acceptable approach, the Commission does not prescribe any particular congestion management method. Order 2000 grants RTOs considerable flexibility to propose a congestion pricing method that is best suited to each RTO’s individual circumstances.¹⁰

¹⁰ PJM RTO Order, page 19.

Learning from Experience

A reluctance to be more prescriptive may have been appropriate when FERC first approved the initial (and at that time, still untested) market designs for each of the currently operating ISOs. But over the last four years, FERC and the operating ISOs have gained valuable, sometimes costly experience in implementing different market designs. Much has been learned from market operations in PJM, New York, New England and California about what works, what doesn't, and why, as well as the shortcomings of trying to foster markets in regions without the market coordination provided by independent regional system operators. Despite continuing experiments, there may be an emerging consensus that certain design features are not only desirable but also essential, while other approaches, once thought to be attractive or simpler, have proved to be more complicated than expected, unworkable and fundamentally flawed.

Several markets have now shown what design theory predicted: simplified uniform and zonal pricing mechanisms cannot provide a workable framework for congestion management, fail to send adequate locational price signals, mask and facilitate the exercise of market power, and require increasingly more intrusive forms of intervention and restrictions to counteract the effects of inefficient price incentives.¹¹ At the same time, three years of planning and two years of unsuccessful pilots have shown that uncoordinated (bilateral) market redispatch cannot get the job done, because the complexities of network interactions require effective coordination through the dispatch arranged by a regional system operator.¹²

In contrast, during the same time frame, coordinated, bid-based spot markets with market-based congestion management have been operating more or less successfully in PJM and New York, applying principles of least-cost security-constrained dispatch, locational marginal pricing and tradable point-to-point financial transmission rights. Faced with the need to reform its market rules and uniform pricing scheme, New England has chosen to adopt a standard design founded on these same principles, and other regions are now considering similar approaches. Within this framework, generators have incentives to follow the dispatch required to maintain reliability, decentralized bilateral trading flourishes in ways that support reliable operations and is supported by open access to the spot market, transmission customers can elect to buy through congestion without cross subsidies, parties are compensated at market prices for providing congestion redispatch services, and investors have at least begun to see forward price signals that encourage market-driven investments at the right locations (though more might be done to improve longer-term price transparency and investment incentives).

This paper draws from these and other experiences a set of foundation principles for a "Standard Market Design." With the experience now available, the lessons learned indicate that FERC

¹¹ Many lessons learned from simplified zonal systems would apply as well to some of the proposals to use flowgate transmission rights, which raise issues similar to those that arise with zonal pricing and inter-zonal congestion rights. For example, the concept of defining "commercially significant flowgates" (CSFs) and offering rights only for these CSFs is analogous to the notion that important and frequent congestion can be predicted so as to define zones and inter-zonal rights, but presumably unimportant and infrequent "intra-zonal" congestion can be managed and priced with less care and no rights. Simplifying assumptions that ignore and fail to price grid realities have been the source of persistent incentive problems in several markets. These issues are discussed further in later sections of this paper.

¹² Congestion Management Working Group of the NERC Interface Committee, "Final Report on the NERC Market Redispatch Pilot," November 29, 1999, filed with FERC on December 1, 1999.

should establish the basic design principles that have worked successfully in PJM, New York and elsewhere as a Standard Design and the default “best practices” approach for other regions. That is not to say that the PJM or NY rules cannot be improved, nor does it mean that FERC should impose on every RTO the entire set of PJM or New York Market Rules, associated software and business protocols, though that shortcut is certainly available for those who choose it.¹³ Rather, there is a set of core principles and essential elements that underlie these markets (and successfully operating markets elsewhere), and these provide the foundation for a Standard Market Design. With this as a proven, workable foundation, important details can then be defined and tailored to any special characteristics of each region. The Standard Market Design could serve as FERC’s best practices benchmark for RTO market design and rule development.

Given the risks of poorly designed markets, arguments in favor of continued experimentation with departures from the core principles should bear a heavy burden of proof. Other innovative approaches can and should still be considered, but the proponents of an untried approach should be required to demonstrate that it is theoretically sound and likely to work, while posing minimal risks if it should fail. Equally important, and consistent with FERC’s directive to meet a “best practices or better” standard, untried approaches should be imposed only if they are likely to improve on the Standard Design principles by more effectively achieving the public interest in an efficient competitive market.

The design principles offered here are neither new nor untried. The fundamentals have been described in numerous technical and policy papers over the last decade or more,¹⁴ and the basic principles underlie the market support requirements of FERC’s Order 2000. Equally important, these principles provide the foundation for successful markets in a dozen regions of the world, including workable markets in the Northeast. A successful, theoretically sound and workable model is available; it simply remains to restate its principles and concepts and apply them broadly.

The Task of Market Design

The policy justification for creating competitive electricity markets is to harness the incentives and dynamics of competition in support of efficient outcomes that enhance the public welfare. Where competition has been successful, it has led to more efficient use of resources, market-driven investments that better allocate risks and rewards, valuable innovation in products and services and more efficient prices. But restructuring a highly complex and long regulated

¹³ Because of their different starting points, not everything in PJM or New York is readily transferable to other regions, and work remains to be done in PJM and New York as well, especially on the mechanisms for inter-regional scheduling and dispatch coordination, inter-regional transmission rights, associated pricing and settlement mechanisms, and a consistent and effective regional approach for assuring supply adequacy.

¹⁴ See, e.g., F.C. Schweppe, M.C. Caramanis, R.D. Tabors, and R.E. Bohn, *Spot Pricing of Electricity*, Kluwer Academic Publisher, Norwell, MA, 1988; Larry Ruff, “Stop Wheeling and Start Dealing: Resolving the Transmission Dilemma,” *Electricity Journal*, June 1994; Don Garber, William Hogan and Larry Ruff, “An Efficient Wholesale Electricity Market: The SDG&E Pooling Proposal,” *Electricity Journal*, September 1994; William Hogan, “Coordination for Competition in an Electricity Market,” Center for Business and Government, Harvard University, March 2, 1995; W. Hogan, “Regional Transmission Organizations: Millennium Order on Designing Market Institutions for Electricity Network Systems,” May, 2000.

industry into one driven by market principles is extremely difficult, and efforts are not always successful. A poorly designed restructuring effort in an industry as complicated and essential as electricity can be disastrous, as reports from California and elsewhere attest. Such failures serve as a warning that successful electricity markets do not happen by chance or neglect, and they may be especially elusive if careful attention is not paid to the fundamentals of market design, market pricing and compatible regulatory support and oversight. Well-designed markets may well be self-correcting, but poorly designed markets and incompatible supporting institutional arrangements are clearly not. Indeed, they can become self-perpetuating.

A key premise of successful competitive markets is that they work through the interaction of private, decentralized trading and investment decisions. An effective electricity market would therefore seem to depend not only on having many buyers and sellers to ensure competitive outcomes, but also on allowing substantial commercial freedom to market buyers, sellers and various types of traders and risk managers. Trading rules would then allow the participants the freedom to fashion and implement various trading and risk management arrangements with each other, at prices to which they mutually agree, in pursuit of their respective commercial interests.

In contrast, the complexities of electricity networks require a degree of centralized coordination over system operations. Since the need for central coordination is always present, the imposition of central coordination would, at first glance, appear to be in conflict with the desire for commercial flexibility and decentralized trading mechanisms. However, this need not be the case in a well-designed market.

The interaction between the need for central coordination to achieve system reliability and the desire for decentralized commercial freedom and flexibility describes the arena for electricity market design. The task is to fashion rules for market operations and pricing that accommodate and reconcile the dictates of reliable operations and the need for commercial flexibility.

Successful electricity markets have confronted and solved this problem by explicitly recognizing the need for coordination, allowing the system operator to coordinate the short-run spot markets associated with maintaining reliability, and then consistently applying market principles and efficient pricing to the necessary coordination services and products.¹⁵ As in any market, getting the pricing principles right is the key. The challenge in designing an electricity market is to define prices that accurately reflect what the system operator must do to maintain reliability – from calling upon more (or less) expensive generators to balance the system to redispatching units out of merit order to relieve congestion -- and thus accurately signal to grid users the true economic consequences of the grid use choices they make.

¹⁵ Ruff, Larry, “Competitive Electricity Markets: One Size *Should* Fit All,” *Electricity Journal*, November 1999, pp. 20-35. As Ruff notes,

Every electricity market in which there has been reasonably effective and efficient competition among multiple sellers and multiple buyers has had a more or less independent system operator (ISO) using some sort of spot market to coordinate the physical operations of the system. Such an integrated dispatch/spot market process is the only practical way to price or internalize the complex network externalities that have traditionally justified vertically integrated monopoly utilities. The resulting spot prices and the sophisticated financial contracting they support create consistency between reliability objectives and commercial objectives.

Efficient price signals consistent with reliability then provide a stage upon which commercial freedom can play. Bilateral contracts and other decentralized trading mechanisms, self scheduling and self commitment can be freely arranged and implemented across a complex, interconnected grid without concern for cross subsidies or conflicts with reliability. Flexible dispatch/scheduling rules can accommodate both coordinated, bid-based trading and bilateral or self-nominated schedules, while flexible net accounting rules can settle transmission schedules, bilateral imbalances and spot trades at internally consistent market prices. The result is to place bilateral and spot trading on an equal footing, leaving the mix between long- and short-run options, and between decentralized trading and coordinated trading, to be decided by the market.

In contrast, unsuccessful markets have failed or performed poorly primarily because they took the opposite course, artificially separating – or attempting to separate – the necessary coordination functions from the market, limiting market access to the system operator’s essential services, and then failing to price the necessary coordination functions at efficient, market-clearing prices. The predictable result has been that both markets and system coordination functions performed badly, requiring substantial intervention in commercial choices, administrative restrictions and penalties (in lieu of market prices) to enforce behavior, and re-centralized investment decision-making to preserve reliability.¹⁶ Explaining the logic behind these observed results and designing RTO markets to avoid the same pitfalls are why a standard market design is most needed.

While faulty market designs can have extremely serious consequences for parties in the affected region, they are not easy to fix. Parties will respond logically to the perverse incentives, and as soon as advantaged parties learn to exploit the inefficiencies, achieving necessary reforms becomes difficult. Market participants advantaged by the rules have no incentive to support (and may actively oppose) market self correction, while regulators, on whom the burden of reform must now fall, are ill prepared to do the job because there is no standard design framework and no coherent set of adopted design principles from which to derive the needed reforms. Faced with pressure to “do something,” regulators may impose *ad hoc* patchwork approaches that not only fail to solve the underlying problems but may make conditions worse or sow the seeds for future problems. Eventually, the need to address flawed concepts in one area requires more comprehensive and lengthy redesign in others, as both New England and California reform efforts have shown.

The common lessons derived from these experiences suggest that FERC needs a standard design framework based on proven, workable approaches and a coherent set of internally consistent market rules.¹⁷ The remainder of this paper offers a suggested framework and then describes the foundation principles for a Standard Market Design for electricity.

¹⁶ See, John D. Chandley, Scott M. Harvey and William W. Hogan, “Electricity Market Reform in California,” Comments in FERC Docket EL00-95-000, November 22, 2000.

¹⁷ State regulators would also benefit from a Standard Design, as well as a compatible set of guidelines for restructuring at the retail level. Incompatible efforts at the wholesale and retail ends of the same market can create huge risks for market participants, affected utilities and consumers. This paper does not address state regulatory issues except to note the obvious need for a complementary state regulatory framework.

System Operations, the Dispatch and Market Design

In electricity the laws of physics dictate certain essential features of system coordination and operations. In the United States, moreover, transmission interconnections create vast, inter-regional systems that must be coordinated and operated under common, consistent procedures and within narrow tolerances. In extreme cases, failure to continuously observe these limits can jeopardize an entire interconnection – up to half the country – in a matter of minutes. Even far less serious system failures can result in costs in the billions of dollars from damaged equipment and lost operations.¹⁸ Some degree of centralized coordination is therefore essential to maintain reliable operations of an electricity system.

The system operator's dispatch plays a pivotal role in integrating short-run markets and reliability. To ensure reliability, each system operator uses the dispatch of generation to maintain system balances at constant frequency and acceptable voltage levels, to control inter-regional flows and to manage congestion within its area of control. These same functions also provide the basis for much of the RTO's market coordination and support. In a market-based structure, the system operator uses voluntary price offers and bids submitted by participants to arrange the dispatch that, along with regulation units, balances the system and manages congestion, while ensuring the availability of operating reserves to deal with various contingencies. The system operator then uses these price offers and bids to define market-clearing prices for the real-time balancing market and market-based congestion redispatch – the same markets that Order 2000 requires every RTO to provide.

In real time, the coordination of system operations to maintain reliability and the coordination of short-run markets for balancing and congestion management converge in an integrated dispatch process, driven by the price offers and bids submitted by market participants for that dispatch. Hence, the starting point for a workable market design is the recognition that real-time markets for balancing and congestion management are based on the system operator's bid-based dispatch.

A workable electricity market therefore requires a multi-function RTO to be both the “system operator” and the short-run “market operator.” Both functions center around the real-time dispatch, driven by participants' voluntary offers and bids and the need to balance the system within its security limits. If the RTO also operates short-run *forward* markets, as the ISOs do in the Northeast, these markets also derive from participant offers and bids that are used to define a day-ahead and/or hour-ahead dispatch that balances the system and honors all grid constraints at the least as-bid cost for that forward period. The RTO can then define market-clearing prices to pay those who offer energy to the dispatch and to charge those who take energy from the dispatch. RTO-coordinated real-time and short-run forward markets arise naturally from the fact that while the ISO is using the offers/bids to arrange a dispatch to achieve reliable operations, the participants are using the bid-based dispatch to buy and sell spot/imbalance energy (and spot transmission) at market prices to support their commercial objectives.

¹⁸ Several now familiar factors dictate the need for coordinated system operations: the speed with which electrical energy travels, the difficulties of storage, and the resulting need for virtually instantaneous balancing between electricity production and consumption; the degree of interconnection across vast geographic regions, and the need for close coordination to avoid the near instantaneous effects of failures in one region on reliability in every other interconnected region; and the absolute physical requirements for maintaining voltage support, constant frequency and stability within extremely close tolerances across the system for both safety and reliability purposes.

The RTO's market coordination functions are thus so closely related to the dispatch/redispach functions of the system operator that it is meaningless to distinguish the "market operator" from the "system operator." The RTO must be both. As poorly-functioning markets have shown, artificially separating these functions, imposing arbitrary limits on participants' ability to use a bid-based dispatch as an open spot market, or restricting the system operator's ability to use market mechanisms to arrange an efficient (least-cost) dispatch for balancing and congestion management, create both reliability problems and market inefficiencies.¹⁹ In well-designed systems, the independent system operator (ISO) and the independent market operator (IMO) are the same entity.²⁰

System/Market Operations and Independence

If an electricity market is to function both efficiently and without discrimination, the entity that coordinates the short-run markets, controls physical access to the grid, dispatches the system and coordinates the provision of essential grid services must do so in an unbiased manner. In the United States and elsewhere, this principle is embodied in the notion of an independent system operator, whose governance and conduct must be free of undue influence from those with commercial interests in market outcomes. If the RTO were not truly independent, participants with commercial interests could exert improper influence and/or obtain preferential treatment to achieve commercial advantages for the sellers, traders or buyers they represent.²¹

The need for an unbiased or "independent" system/market operator applies no matter how the mechanisms for grid ownership are structured. An RTO formed around a single "transco" needs its system/market operator to be independent for exactly the same reasons as an RTO formed around several "gridcos," and the need does not change whether the grid ownership is structured to be "for profit" or "non-profit." No matter who owns the grid, or how the grid owners/investors recover their investments, the RTO's system/market operator must be impartial, independent from market participants, and must perform the essential coordination functions in an unbiased manner. To ensure this impartiality, the system/market operator may need to be prohibited from engaging in any commercial activities that might affect its ability or incentives to perform its functions impartially. Because market coordination and operations can affect the value of any infrastructure investments and thus materially affect the incentives for market-driven investments, those entities eligible to undertake such investment should be separated, or at least walled off, from the entity that performs the ISO/IMO functions.

Efficiency and Economic Dispatch

Each RTO must perform its essential dispatch and related market coordination functions in an efficient manner. This goal is best exemplified by the familiar principles of a security-

¹⁹ Hogan, William, "A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy," *The Electricity Journal*, December 1995, pp. 26-37.

²⁰ The new Ontario market makes this identity explicit by calling the system operator the "independent market operator" (IMO). However, the same concept lies at the foundation of every working electricity market, no matter what the coordinating entity is called.

²¹ Allowing an important buyer to gain preferential access to an RTO's system and market coordination activities is no more acceptable than allowing preferential access to selected sellers or traders. It is equivalent to allowing a vertically integrated utility to use the dispatch preferentially to advantage its preferred customers or its own generation and contracting activities.

constrained economic dispatch. Security-constrained economic dispatch is common to most systems and is well understood by system operators, whether they work for an ISO/IMO or utility-run control centers. To achieve an economic dispatch, the system operator uses each plant's operational and economic inputs to arrange a dispatch that simultaneously balances the system and meets all security constraints, and does so at the lowest economic cost, given the cost parameters available to the system operator.²²

In a system in which generation is independently owned, and the market is coordinated by an independent system/market operator rather than a utility system operator, the RTO obtains the necessary dispatch inputs in the form of voluntary generator offers and load bids. Parties that do not wish to be subject to the system operator's dispatch can self schedule (or bid extreme prices that assure their dispatch preference), and they become price-takers. For those who choose to participate, the offers and bids indicate each party's willingness to sell or buy different quantities of energy at various prices through the dispatch during each dispatch interval. The RTO can then use these indications of market preferences to do what it must do for reliability – arrange a dispatch to balance the system, manage the flows between regions, relieve congestion and honor all security constraints – all at the lowest as-bid cost. By defining market-clearing prices for the energy offered and provided through the dispatch, the RTO can offer the participants an efficient means to buy and sell energy to support a wide range of commercial transactions – from decentralized, bilateral transactions to coordinated spot transactions -- while providing efficient price signals that are consistent with short-run operations.

Ensuring an economic dispatch is critical because the dispatch plays such a central role in supporting an efficient market. In addition to providing essential balancing and constraint management, the dispatch is the physical “provider of last resort” in all electricity markets.²³ No matter what the load is, or how many customers have selected alternative providers, the dispatch ensures that all loads are served, so it is essential that they be served at the lowest as-bid cost. In addition, the real-time spot market that flows from this dispatch provides the support for other market arrangements. For example, real-time spot prices (or day-ahead prices, if there is an RTO-coordinated day-ahead market) provide a reference for writing forward contracts. If the dispatch upon which spot prices are based is not economic, the higher costs are likely to be reflected in forward contracts as well. The spot prices also effectively eliminate the problem of liquidated damages when either contracting party fails to perform (i.e., to generate or consume) as expected. It does not matter whether the supplier injects more or less energy, or the buyer withdraws more or less energy than anticipated by the contract schedules, because the resulting

²² FERC encourages RTOs to develop incentives to promote more efficient grid operations. The real-time dispatch is arguably an RTO's most important function. Rules that allow and encourage an RTO to arrange and implement this dispatch as efficiently as possible are necessary to achieve the goal of efficient grid operations. Indeed, it is hard to imagine how an RTO could achieve efficient grid operations without an efficient security-constrained economic dispatch. Once again, the RTO dispatch and the manner in which it is arranged and priced are fundamental to an RTO's success.

²³ The function of the “financial” provider of last resort could be met by some designated entity or entities that would be responsible for managing price risks on behalf of uncontracted (“default”) customers, implying a continuing regulatory oversight of such entities. Alternatively, this function could be met – or eliminated -- by passing the RTO's spot prices through to the end-use consumers, *provided* they have effective and readily available means to hedge spot price volatility, a condition not likely to be fully satisfied when new markets open. Eventually, however, consumers who wish to avoid the uncertainty of spot prices can arrange their own risk management by choosing to enter into contracts with any retailer or entity that can provide risk management services. Under this arrangement, no official provider of last resort need be specified.

deviations or imbalances are simply met or absorbed through the dispatch as it automatically balances the system, and these imbalances can then be easily settled at the spot prices.

Open access to the RTO's imbalance/spot market also allows contracting parties to avoid the burden and expense of precise or even approximate load following. Aggregate system balancing is a reliability necessity, and that is what the dispatch provides. But since the dispatch is open to all users, mandating individually balanced schedules is not needed and may not even be practical in many cases, let alone commercially desirable. Parties are free to match their schedules if they choose, but if they can't, or choose not to, each schedule's imbalances are simply supplied or absorbed in the aggregate by the RTO's real-time dispatch, with the implicit purchases and sales by each party settled at the market-clearing spot prices that flow from that dispatch. With an open spot market, moreover, generators have a ready market for their uncontracted output, and loads and retailers have a reliable source and dependable market to obtain energy to meet their uncontracted demand or cover their contract obligations. An open spot market thus facilitates entry by new suppliers and retailers.

Restrictions on access to the dispatch-based spot market are therefore as misguided as restrictions on forward contracting; the options are complementary and both are essential pieces of a workable market. If the dispatch/spot prices are efficient, arbitrary percentage limits on spot transactions versus forward transactions are not needed and only serve to restrict commercial freedom and raise overall market costs. The key is to get the pricing right and let market choices and commercial needs define the appropriate mix between forward contracts and coordinated spot trading through the RTO's dispatch/spot market.

A Regional Spot Market Requires A Regional Dispatch

A regional spot market to support regional trading requires a regional dispatch. Most regions outside the Northeast (and California) are currently operated through smaller local control areas that do not allow open access to their dispatches nor market-priced spot markets. In these regions, important transition steps must be taken to move from local control area dispatches to a regional dispatch that is coordinated and priced by the RTO as an open spot market and effective tool for regional congestion management. Initially, this may require a hierarchical approach in which the RTO implements its regionally coordinated dispatch through existing local control areas.

New RTOs must eventually go through this dispatch integration process, because the reach of a coordinated dispatch largely defines the scope for regional balancing, congestion and loop-flow management, consistent pricing and transmission rights. The broader the scope of this dispatch, the broader the reach of the regional market.

With a regionally coordinated dispatch as the foundation, an RTO can define related market rules through consistent application of economic and market principles and careful observance of the requirements for reliability. Bid-based, RTO-coordinated markets can be designed to be open to all participants on a non-discriminatory basis. The RTO can define market-clearing prices to settle all purchases and sales in its coordinated markets, leaving participants free to pursue a wide range of commercial transactions and choices, with price signals consistent with the RTO's reliability mandates. These features define the essential components for RTO market support, as required by Order 2000, and provide the foundation principles for a best practices Standard Market Design.

Foundational Principles for a Standard Market Design

- Each RTO must have a system operator that coordinates system operations through a regional, security-constrained, economic dispatch. The RTO uses this regional dispatch to maintain system balances and adjusts the dispatch to relieve congestion and keep flows within all security limits, at the lowest as-bid cost. RTO coordination of the region's dispatch is therefore the cornerstone for RTO market support.
- An RTO would use market mechanisms to support its essential system operation functions, first by accepting price offers and bids from those generators and loads eligible to participate in the RTO's security-constrained economic dispatch, and second by defining market-clearing prices for the energy bought and sold through the dispatch. The RTO would thus operate an open, bid-based spot market.
- As its market coordination capabilities mature, an RTO would use similar market processes to coordinate the acquisition and deployment of resources for regulation and operating reserves. The RTO would coordinate these markets with its real-time energy market to achieve consistent, market-clearing prices across all of its coordinated short-run markets. RTO-coordinated forward markets (e.g., day-ahead) could be added, if desired by participants, to provide further price certainty for energy and transmission prices and a means to deal with startup/commitment risks.
- Market participants – generators, loads and traders – would participate in the provision and pricing of these services through the voluntary submission of quantity/price offers and bids to the RTO, for its use in system balancing and spot trading, constraint/congestion management, and maintaining adequate regulation and reserves. Under a consistent pricing and settlement framework to prevent cost-shifting, “self-provision” of these latter services would also be accommodated.
- The RTO would function as both an independent system operator and an independent market operator. RTO operational and market coordination functions that center around the dispatch are inseparable in real time.
- To ensure unbiased (non-discriminatory) and efficient operations, these integrated functions would be performed by an entity with no commercial interest in market outcomes. Operating an open, efficient market in an unbiased fashion would be regarded and protected as an activity with public interest objectives. These involve operating fair, open, transparent and efficient competitive markets for purposes of promoting economic welfare and an efficient allocation of resources, while ensuring reliable system operations. RTO activities, rules, procedures and affiliations inconsistent with these public interest objectives would not be permitted.

These principles provide the initial foundation for the market support functions each RTO must provide. In addition, FERC's principles of open access, driven by the twin policy objectives of economic efficiency and non-discrimination, require additional foundation principles for a Standard Market Design:

- Market participants would have open, non-discriminatory access to the essential products, services and markets coordinated by the RTO. For example, there would

be no arbitrary limits on the ability of parties to participate in the RTO's bid-based dispatch. There would be no limits on their right to use the resulting balancing and congestion management markets that flow from the dispatch coordinated by the RTO.²⁴ Open access to balancing markets means an open spot market for buying and selling energy at market-clearing prices.

- Whatever services the RTO must provide, it should provide efficiently. For example, there should be no restrictions on the RTO's ability to arrange an economic dispatch – i.e., at the lowest-as-bid cost -- given the price offers and bids submitted by the participants and the security constraints that the RTO must honor in the dispatch.
- Whatever services the RTO must provide, it should price efficiently, using (wherever practical) market-clearing prices derived from market bids and principles of marginal cost pricing.
- Consistent with efficient pricing, the RTO would price the dispatch (and congestion redispatch) as accurately as possible, so that the resulting price signals reflect the effects of congestion and are consistent with what the ISO requires parties to do to relieve congestion and maintain reliability.

These last principles require that market prices be based on principles of marginal costs and be consistent with security-constrained economic dispatch. Prices consistent with the dispatch would therefore reflect the marginal cost of any redispatch required to relieve congestion and accommodate each transaction. These are the principles that underlie nodal spot pricing, or locational marginal pricing.²⁵ An RTO would apply the principles of locational marginal pricing to define market prices for spot energy and transmission.

- The price for imbalance/spot energy at each location on the grid – the “nodal price” -- would be the locational marginal price, which is defined as the incremental cost in the dispatch of meeting an increment of load at each location, given the price offers and bids,²⁶ the actual dispatch, and the grid constraints affecting that dispatch.
- Generators providing energy for the dispatch would be credited in the RTO settlements for their injections at the nodal prices for their respective locations. Additional injections net of bilateral schedules would therefore be sales to the spot market and would be paid the nodal prices.
- Loads served by the dispatch would be debited in the RTO settlements for their withdrawals at the nodal prices for their respective locations. Additional

²⁴ Order 2000 recognizes that open access to the real-time balancing market is required to ensure non-discriminatory access to transmission. See, Order 2000 at p. 425.

²⁵ Nodal prices would differ by location because of the effects of congestion and losses. An initial pricing system might not reflect marginal losses (as in PJM), but complete nodal systems would (as in New York).

²⁶ In calculating the market-clearing prices from the dispatch, the RTO would use the offers and bids of parties that were actually following the RTO's dispatch instructions within some defined tolerance. All other parties would be viewed as setting their own schedules and would be settled as price-takers.

withdrawals net of bilateral schedules would therefore be purchases from the spot market and would be charged at the nodal prices.²⁷

- Schedules for bilateral contracts would be fully accommodated by the RTO settlements. Imbalances (deviations) from bilateral schedules would be settled at the nodal prices, such that
 - (1) generators injecting less (more) than their schedules would be charged (paid) the nodal price at their location for their deviations and
 - (2) loads withdrawing more (less) than their schedules would be charged (paid) the nodal price at their location for their deviations.
- The price for transmission usage for any transmission schedule (bilateral transaction) would be the marginal cost of the redispatch necessary to accommodate that schedule or transaction. The marginal redispatch cost for each megawatt of a schedule or transaction is equal to the nodal price at the point of delivery minus the nodal price at the point of receipt.²⁸ The usage charges that apply to transactions that are “scheduled” with the RTO can be applied as well as to transactions with “loop flows” across the RTO-controlled grid from transactions scheduled outside the RTO.²⁹

Efficient Pricing and Market Flexibility

Order 2000 requires that an RTO’s market mechanisms provide all grid users with efficient price signals that reflect congestion and expansion costs and give participants efficient incentives regarding the consequences of their transmission usage.³⁰ Providing efficient signals regarding short-run usage and long-run investments is the key to giving participants commercial flexibility in choosing and implementing transactions. Where these pricing rules are followed, generators and other parties can be given substantial commercial freedom to respond to the resulting price signals without significant concern that their commercial decisions will undermine reliability. Where these rules are not followed, the system operator must inevitably intervene in the market through administrative curtailments, restrictions and non-market penalties.

²⁷ In most regions, small customers do not have interval meters, so applying hourly (or shorter) nodal prices requires some averaging and the use of load profiles. Also, such loads may not yet be accurately mapped to individual buses (nodes). In these cases, loads (or Load-Serving Entities purchasing on their behalf) can be charged an aggregate nodal price, defined as the load-weighted average of the nodal prices in the area in which the loads are located. For state retail purposes, the aggregations can cover a service area or portions thereof for some transition period.

²⁸ For simplicity, this difference is also referred to as the “nodal price difference” or “locational price difference.” The difference may be positive or negative.

²⁹ An RTO that uses this market-based method to price the redispatch costs imposed by such loopflows avoids the dilemma of having either to absorb uncompensated redispatch costs or to impose TLR curtailments as its only alternatives. Instead, the RTO provides a market-priced redispatch service and allocates grid use efficiently to those willing to pay the marginal costs of that redispatch.

³⁰ Order 2000 at pp. 382, 489.

Ample experience in markets from New England to California demonstrates the importance of applying principles of marginal cost and aligning real-time market prices with the security-constrained economic dispatch used to maintain reliability. If imbalance/spot market settlement prices are not allowed to reflect actual market-clearing prices derived from marginal costs, parties have incentives to use the imbalance/spot energy markets in ways that shift costs to other parties. Similarly, if transmission usage charges do not reflect the full opportunity costs of any usage (that is, the marginal costs of any redispatch required to accommodate a transmission schedule), parties have incentives to over-schedule transmission usage. Accommodating these schedules through redispatch would then shift redispatch costs to others and encourage more parties to over schedule, providing an incentive to increase congestion.

Where an RTO's prices do not efficiently allocate access to the grid, the RTO must use administrative restrictions to limit access to keep flows within security limits. That is why control areas that provide no mechanism to price the marginal cost of redispatch must resort to physical curtailments under TLR. For similar reasons, an RTO using inefficient pricing would also have to limit access to its imbalance/spot energy market.³¹ Balanced schedule requirements and imbalance penalties (several RTOs), penalties to discourage "under-scheduling" (California), denial of "congestion buy-through" (everywhere but PJM and New York), restricted access to avoid curtailments (throughout the West except California) and curtailments under TLR (throughout the Eastern Interconnection, except PJM and NY) are all real-world examples of administrative restrictions that have been (and continue to be) used to offset the incentives of inefficient pricing rules.

Failure to use market-clearing prices that reflect marginal costs at each location also requires an RTO to use non-market side payments to persuade plants required for congestion redispatch to follow dispatch instructions. Side payments to constrained-on and constrained-off generators are needed because uniform or zonal settlement prices, which are not based on locational marginal costs, are inconsistent with participant bids and the requirements for redispatch. Constrained-off generators must be paid not to run, or else administrative controls must be placed on plant operations to prevent them from operating in response to the inefficient price signals provided by a uniform or zonal price. However, experiences in several markets show that these side payments encourage strategic bidding to maximize the side payments.³²

³¹ Experience with non-locational pricing in PJM during 1997 showed that where pricing for grid usage is inefficient, restrictions on access to the grid must be coupled with restrictions on access to the balancing/spot market. The reason is that selling energy at location A and buying energy at location B through the imbalance/spot market is equivalent to scheduling transmission from A to B. If parties are limited through one mechanism, they will exploit any equivalent mechanism in response to the same incentives. In PJM, parties implementing internal transactions during 1997 were allowed access to the spot market to support their transactions, while parties implementing exports and imports were denied access to the spot market, to avoid having the redispatch costs imposed on the PJM members. The PJM rules did not allow open access to the spot market for all parties until the ISO implemented nodal pricing and was able to settle each transaction using the nodal price that applied to its locations.

³² Such side payments have been used, and subsequently abused, in the United Kingdom, PJM (1997), California (today) and ISO-NE – that is, in markets that do not use nodal locational pricing. The constrained-on and constrained-off bidding games first arose in the UK and were later observed in extreme forms in the California market, which now sets a floor on bids to reduce gaming incentives. PJM eliminated the problem in 1998 by moving to nodal pricing, which does not require such side payments. New England avoided some of these problems by not compensating constrained-off plants, while limiting self scheduling. New England now plans to implement nodal pricing.

In some cases, inefficient pricing rules may also encourage parties to schedule during forward periods in ways that create artificial congestion. The RTO must then pay the same parties to relieve this congestion in real time.³³

Administrative restrictions on investment decisions are also required, because the prices fail to provide efficient investment signals. If real-time prices are inefficient, then expectations of forward prices will not provide accurate signals for future investments. For example, if real-time and forward market price signals do not clearly indicate whether new generation at a location will help solve congestion or make it worse, the RTO must intervene in new investment and siting decisions, imposing interconnection restrictions and upgrade costs on new entrants, and so on. FERC has recognized that these restrictions discourage new entry and thus hinder the development of more competitive supplies. In both California and New England, FERC rejected such restrictions and directed the ISOs to reform their congestion management and pricing rules.³⁴

The universal lesson from many markets is that inefficient pricing schemes create incentives for short-run strategic behavior while encouraging participant decisions that are inconsistent with reliability, thus forcing intervention by the system operator. This necessary intervention directly limits market freedom and flexibility. Because no system operator can allow market actions to jeopardize system reliability, the system operator must eventually intervene in the market and impose administrative controls and non-market penalties to maintain reliability, discourage gaming and limit cost shifting.

³³ This response to poor pricing signals arises in California and is partly a function of using unrealistic “commercial models” to manage only inter-zonal congestion in the forward markets, while the ISO must use more realistic operational models to manage all of the congestion in real time. The same concept of managing only “commercial” congestion in the forward market and dealing with “operational” congestion in real time can be seen in some of the RTO “hybrid” market design proposals. The concept has been a source of persistent gaming and operational problems in California and should be avoided. See, e.g., FERC’s “Order Accepting for Filing in Part and Rejecting in Part Proposed Tariff Amendment and Directing Reevaluation of Approach to Addressing Intra-zonal Congestion,” January 7, 2000. In that order, FERC noted that the California scheme for zonal pricing and intra-zonal congestion management was “fundamentally flawed” and required “a comprehensive replacement congestion management approach.”

³⁴ For California, see FERC’s “Order Rejecting Proposed Tariff Revisions,” in Docket ER99-3339-000, issued September 15, 1999. For New England, see FERC’s “Order Conditionally Accepting Compliance Filing, as Modified, and Accepting, in Part, and Rejecting, in Part, Proposed Tariff Changes, as Modified,” in Docket No. ER98-3853-000, issued October 29, 1998; also see, “Order Extending Congestion Cost Allocation Methodology on ISO-NE’s Proposal to Replace New England’s Uniform Pricing on an Interim Basis and Deferring other Issues,” in Docket Nos. EL00-59-000 and ER00-2005-000, issued June 1, 2000. In the latter Order, FERC approved ISO-NE’s proposal to retain a scheme that required that average redispatch costs be charged to parties through an uplift until it could be replaced with a system based on locational marginal pricing. In the interim, however, FERC noted at pages 6-7:

Although we will retain the current allocation methodology for now, we will not permit socialization of congestion costs to continue indefinitely. Given the increase in congestion in New England and the significant planned generation addition, it is essential that the ISO implement a new CMS [congestion management system] that relies on market mechanisms to establish price signals that will serve to allocate constrained transmission to the highest valued users and give generation an incentive to locate in appropriate areas.

On the other hand, where energy and transmission prices accurately reflect the marginal costs of congestion redispatch, the settlement prices encourage generators to follow this dispatch. By eliminating the need for side payment schemes, nodal pricing removes the opportunity to gain commercially from strategic bidding to maximize the payments. The resulting market-clearing prices tend to reinforce reliability, rather than undermine it. Short-run market prices then provide an appropriate foundation for forward price signals for siting new generation (and loads) at locations that will help relieve congestion, rather than exacerbate it, while providing an incentive for market-driven transmission upgrades.

Efficient pricing is therefore another cornerstone in the Standard Market Design. Achieving commercial flexibility and liberating decentralized decision-making requires that an RTO price its essential coordination functions efficiently, using principles of marginal costs, and in a manner consistent with the dispatch it implements to maintain reliability. With an efficient pricing foundation that supports reliability, the Standard Design can then accommodate a broad range of commercial flexibility and market trading options. For example:

- Market rules can offer market participants maximum commercial freedom with respect to real-time operations, forward trading, and long-run investment decisions. The RTO can support such commercial flexibility without significant concerns for reliability, unfair cost-shifting or inappropriate investment decisions that have either short-run or long-run adverse implications for system security. Within a nodal pricing framework:
 - Parties may choose to participate in the RTO's coordinated regional dispatch. Generators may submit voluntary quantity/price offers to the RTO dispatch, and loads may submit voluntary quantity/price bids to the RTO dispatch.
 - Any participant may buy and sell spot energy through the RTO real-time dispatch/imbalance market and be settled at the nodal spot prices without concern that such transactions will impose unfair cost shifts or subsidies on others. Spot purchases and sales defined by marginal costs do not "lean" on the RTO or anyone else and therefore need not be discouraged, limited or penalized.
 - Generators and Load-Serving Entities may implement bilateral arrangements through either scheduling or settlements.³⁵
 - Parties may submit balanced schedules or not, depending on their commercial needs.³⁶

³⁵ The distinction here means that bilateral parties do not have to "schedule" their bilaterals in advance, though they may choose that option. Alternatively, they can self-schedule individually, or participate in the dispatch individually, and then inform the RTO of their bilateral arrangement only for settlement purposes, after the dispatch is completed. In the latter case, the RTO settlement nets the supplier's credits (its injections times the nodal price at its location) and the buyer's debits (its withdrawals times the nodal price at its location) to produce a net bill that reflects imbalances and any congestion-related usage charge.

³⁶ The freedom for both generators and loads to submit offers and bids to the RTO for its use in system balancing and congestion redispatch without regard to balanced schedules adds to the depth and competitiveness of the real-time balancing and congestion management markets. The more competitive these markets are, the more effective they are in managing the system through market prices and voluntary market choices, rather than RTO administrative decisions. Allowing loads to bid without balanced

- Parties may offer non-contracted energy to the RTO for dispatch, purchase spot energy to support a bilateral, and use the balancing/spot market to purchase energy to meet any uncontracted demand. Such purchases and sales are settled at the market-clearing spot prices.
- Any transmission customer (i.e., those with bilateral schedules) may submit transmission schedules that the RTO will accommodate through bid-based economic redispatch arranged by the RTO, as long as redispatch bids are available and the customer agrees to pay the marginal cost of redispatch to accommodate the transaction. A party may thus choose whether to “buy through congestion,” based on its willingness to pay the transmission usage charge, or choose to self-curtail if it expects the transmission usage charge (equal to the nodal price difference) to exceed its willingness to pay.³⁷ Parties otherwise subject to TLR curtailments due to loopflow through the RTO system have the same options.
- A transmission customer may implement a schedule that provides “counter-flow,” which lowers the marginal cost of redispatch, and be compensated accordingly for the value of that counter-flow.³⁸ Parties thus have another means to provide congestion management and be paid its market value.
- Any Load-Serving Entity (LSE) may self-schedule its own generation to serve its own loads and/or use the RTO balancing/spot market in any mix it chooses. The LSE is subject to the same imbalance/spot pricing and transmission usage charges as any other party. There is no bias in favor of or against “native loads.”
- Any generator, trader, customer or LSE may determine the mix of contract and spot purchases or sales appropriate for its commercial circumstances.
- The RTO can create financial trading “hubs” based on any aggregation of nodes (grid locations where spot prices are defined by the RTO using nodal pricing), with the hub prices defined as the fixed-weighted average of the prices at the nodes included in each hub. Market participants may then freely trade at, to and from these hubs and have their energy trades and transmission rights settled by the RTO using the hub prices. Similarly, retailers can have any collection of nodal prices aggregated to facilitate settlements with loads at multiple locations.

schedules also provides an effective means for defining market-clearing prices during shortage conditions and mitigating the potential for market power.

³⁷ Within the nodal pricing framework, the opportunity cost of any transmission usage equals the marginal cost of redispatching the system to accommodate that usage, which is in turn equal to the difference in nodal prices. The RTO would continuously post the nodal prices in near real time to allow parties to exercise these options more effectively.

³⁸ The value of the counterflow is defined by the savings in opportunity costs or marginal redispatch costs, as defined by the nodal price difference that applies to the points of receipt and delivery for the counter-flow schedule.

- Generator investment and siting decisions can be market-driven without significant RTO direction or interference.
- New generator interconnection procedures can focus on procedures to ensure safe and reliable interconnection and not become mechanisms to restrict new entry for fear of impacts on grid congestion. Nodal prices and related congestion (transmission usage) charges, and expected forward prices based on these real-time prices, will tend to discourage siting at locations that worsen congestion and tend to encourage siting at locations that relieve congestion. Interconnection rules therefore need not discourage new entry by imposing system (“deep”) upgrade costs as a condition for new interconnections.
- Through the award of transmission rights to those who fund upgrades and increase system capacity, at least some transmission investments can be market-driven, in the absence of market failure, and should be accommodated by the RTO.

Transmission Rights in a Market with Nodal Pricing

Under the Standard Market Design, the risks associated with uncertain physical delivery due to congestion are virtually eliminated by the RTO’s dispatch. Whenever proposed schedules would create congestion, the RTO would provide redispatch for parties willing to pay the marginal costs of any redispatch needed to accommodate those transmission schedules, rather than ration grid access through physical curtailments based on administrative priorities. By eliminating the risks of physical delivery, the availability of an RTO redispatch option allows market participants to focus on the financial risks associated with their transactions.

In the Standard Market Design, the financial risks associated with congestion become transparent through the locational marginal (nodal) prices and associated usage charges, but parties need an appropriate financial instrument to hedge these risks. The RTO will use nodal prices to settle imbalance/spot energy purchases and sales; it will use nodal price differences between points of receipt and delivery to define transmission usage charges that reflect the marginal cost of redispatching the system to accommodate each transaction. However, parties will not know the effect of congestion on spot energy prices and transmission usage charges until after they complete their transactions and after the RTO implements the dispatch and calculates the nodal prices at each location. Spot pricing is *ex post*.

Participants will therefore need a system of transmission rights to manage the risks of price uncertainty arising from pricing congestion. This purpose can be expressed in different ways:

- Transmission rights in a nodal pricing system provide a mechanism by which a party can lock in the price of transmission – the usage charge -- in advance.
- Transmission rights in a nodal pricing system allow a party to get access across the grid to the price at another location, even though there may be congestion between that location and the party’s location.³⁹

³⁹ In a nodal pricing system, a “location” for settlement purposes can be defined as a specific bus or node or as the weighted average of many nodes. For a Load-Serving Entity, for example, its location can be defined by all the nodes serving that load, and this aggregated settlement price can be defined as the load-

- Transmission rights in a nodal pricing system provide a means by which a party can offset the impact of congestion on prices at any location.

An effective transmission right (or set of rights) for a transaction from A to B would allow a party implementing that transaction to be credited for the same nodal price difference that defines the usage charge for that transaction, so that the credit received for a right matching a transaction would exactly offset or hedge the usage charge for that transaction. A party holding that transmission right would be perfectly hedged against any A-to-B usage charge (the congestion-related differences in nodal prices) and would thus have effective access to the price at location A, the generator/source location, even during congested hours, as the following illustration shows.

Transmission Rights as Financial Hedges	
Where the nodal price at generator/receipt location	= Nodal Price at A
And the nodal price at load/delivery location	= Nodal Price at B,
And the party holds a right from A to B	
Then,	
The transmission usage charge	= Nodal Price at B minus Nodal Price at A
The transmission right credit	= Nodal Price at B minus Nodal Price at A
The transmission right serves to offset or “hedge” the usage (congestion) charge.	
(Note that the hedge in this illustration functions whether the Nodal Price at A is higher or lower than the Nodal Price at B, because the usage charge will also change in an exactly corresponding fashion. See discussion of “obligations.”)	

In this illustration, the transmission rights are directional and defined from point to point (location A to location B). The RTO defines the settlement credit for the financial transmission rights using the same nodal prices that it uses to define the transmission usage charges.⁴⁰ This symmetry makes the operation of the rights transparent and intuitive, while simplifying the settlement system.⁴¹

weighted average of the nodal prices that make up that aggregate. Any trading hub can be defined in a similar manner.

⁴⁰ Revenues to fund these hedges would come from the settlement surplus collected by the RTO in its coordinated markets. The surplus arises naturally in a system that prices congestion and derives here from usage charges (differences in nodal prices) paid by transmission customers and from the fact that the aggregate revenue from spot energy prices paid by loads is greater than the payments made to generators.

⁴¹ In theory, an equivalent credit could be defined consistent with the same nodal prices by pricing each of the binding constraints between the source and sink locations. The credit for a matching set of such financial “flowgate” rights for the exactly equivalent set of constraints would *in theory* provide the same hedging value in a nodal pricing system as the point-to-point right assumed in the illustration. However, equivalency under all conditions depends on assumptions that the binding constraints are predictable and stable, and that the flows across them are predictable and stable. If these assumptions do not hold, then the equivalency disappears. Where RTOs are considering the use of flowgate rights, special care must be taken to ensure that settlement rules do not assume equivalent credit under conditions that do not support

In addition, the settlement value of a financial transmission right between any other set of locations – e.g. a right between locations C and D -- would be defined in the same manner, as the difference in the nodal prices at those two locations. Thus, any financial transmission right (or set of rights) could be used to hedge (or partially hedge) any schedule's usage charge, and there would be no requirement that the rights match the schedules. Parties could therefore select, acquire and trade rights for risk management purposes independently from the implementation of, or changes in, their actual schedules or the RTO's economic dispatch (or congestion redispatch), and the rights would still serve the risk management function.

Order 2000 Requirements for Transmission Rights

FERC's Order 2000 requires that each RTO provide a system of tradable transmission rights that can hedge locational price differences resulting from congestion. Further, FERC requires that the transmission rights system must promote an efficient dispatch.⁴² This implies that the system of transmission rights should not encourage or lock the rights holders or the RTO into an inefficient dispatch.

Nodal pricing is premised on an efficient dispatch. The RTO uses voluntary price offers and bids to arrange a security-constrained, economic dispatch. This dispatch balances the system while honoring all constraints at the lowest-as bid costs. This is consistent with the FERC directive that generators dispatched in the presence of congestion be those that can serve loads at the lowest cost.⁴³ Hence, the system of transmission rights for a Standard Market Design must be compatible with, and not undermine, the RTO's security-constrained economic dispatch.

Order 2000 also requires that the RTO's prices provide efficient signals regarding short-run operations and long-run investments. Under Order 2000:

[E]very RTO must establish a system of congestion management that establishes clear rights to transmission facilities and provides market participants with price signals that reflect congestion and expansion costs. (p. 489)

“[W]e will require the RTO to implement a market mechanism that provides all transmission customers with efficient price signals regarding the consequences of their transmission use decisions. We are convinced that efficient congestion management requires that transmission customers be made aware of the cost consequences of their actions in an accurate and timely manner, and we believe that this is best accomplished through such a market mechanism.” (p. 382)

FERC therefore requires a system of transmission rights to do more than just hedge congestion charges. The transmission rights must also support an efficient dispatch and support a pricing system that provides efficient price signals that accurately reflect congestion and expansion costs. The combined system must efficiently allocate grid use to those who value it the most.

equivalency, to avoid creating inefficient arbitrage opportunities and undermining the efficient incentive properties of the nodal prices. Principles that should guide an RTO on this question are discussed below.

⁴² Order 2000, at p. 333.

⁴³ Order 2000, at pp. 332-333.

Financial Transmission Rights in Practice

PJM and New York each provide a system of point-to-point financial transmission rights, or FTRs,⁴⁴ that function like those in the illustration above. These FTRs entitle the holder to a settlement credit equal to the difference in nodal prices (or the congestion component, where the nodal prices also include marginal losses), thus providing a means to hedge congestion-based usage charges that are defined by the same nodal price differences. Moreover, because these FTRs are financial, they do not interfere with or undermine the ISO's least-cost dispatch. This financial aspect can be expressed in different ways:

- The rights entitle the holder to a monetary credit in the RTO settlements; they do not guarantee physical access to the grid.
- Parties are not required to acquire or hold FTRs as a condition for scheduling or gaining grid access.
- Parties are not required to match their FTRs to their schedule, or to match their schedules to their FTRs.
- Parties are free to participate or not in the ISO's security-constrained economic dispatch, whether or not they hold FTRs, and the dispatch does not depend on the FTRs they hold.
- In the RTO settlements, parties receive the market value of the FTRs they hold (defined by the nodal price differences), whether or not they implement a matching schedule.

These financial attributes of FTRs help ensure that parties receive the risk management value of the rights they hold while remaining free to adjust the points of delivery and/or receipt for their schedules or participate in an efficient dispatch irrespective of the rights they hold. Parties are not discouraged from making these adjustments and have no incentive to undermine an efficient dispatch, whether or not they hold rights, and the RTO is free to arrange an efficient dispatch or redispatch without having to worry about whether the resulting flows match each party's rights. This would not be the case, however, if the rights were "physical."

If the rights were "physical," parties would need to acquire rights matching their expected transactions as a condition for gaining grid access. If they changed the points of receipt or delivery for their expected transactions, they would have to exchange the rights they held for others that matched their revised schedules. If they were unsuccessful, they would be excluded from the grid and/or risk losing the value of any rights that no longer matched their expected transactions.

⁴⁴ This paper uses the generic term FTR to mean "financial transmission right." FTRs are called "fixed transmission rights" in PJM and "transmission congestion contracts" in New York. While they are allocated differently, in all fundamental respects, PJM's FTRs and New York's TCCs are identical. An identical approach is under development for ISO-New England, where the rights are called "financial congestion hedges." A Northeast RTO would probably use the generic term "financial transmission right."

Parties that held physical rights could also exclude others from using the grid merely by holding these rights until scheduling deadlines were passed. “Use-it-or-lose-it rules” have been proposed to prevent such hoarding in proposed physical rights regimes. However, if such rules were strictly enforced, they would limit trading flexibility, potentially forcing parties to trade rights they might need later before they were certain about this need. If such rights were traded on a recallable basis to avoid losing them, and then recalled by the original party, the counter party would be left with an unreliable right and little time to hedge the risks created for its schedules.

Given these risks, parties holding physical rights would be discouraged from changing their expected points of receipt and delivery, even if it meant a more profitable transaction or a more efficient dispatch. Such changes would require parties to engage in additional trades to realign their rights portfolio with their revised transactions. At any given moment, the distribution of rights would essentially define the dispatch, making it difficult for the RTO to arrange an efficient dispatch that could require generators and rights holders to reallocate the necessary rights to accommodate the more efficient dispatch. Furthermore, an efficient redispatch to relieve unexpected congestion would be almost impossible to achieve in real time, unless the RTO simply ignored the rights.

Physical rights systems applied to an electricity network thus interfere with an efficient dispatch and efficient congestion management coordinated by an RTO. They tend to discourage more efficient operations by the RTO, while imposing additional trading costs and risks on the market that are difficult to manage.

In contrast, financial transmission rights avoid these dilemmas because they are financial instruments that can be acquired and traded independently from the physical dispatch of energy. Because financial rights do not control physical operations, retaining their value need not compromise efficient physical outcomes nor complicate the physical redispatch required to ensure reliability when the grid is constrained.

Consistency Between Financial Rights and Nodal Prices

An attractive and essential feature of the point-to-point FTRs offered by PJM and New York is that each ISO defines an FTR’s settlement value using the same nodal prices used to define congestion usage charges. As shown in the illustration above, the settlement prices for FTRs and the settlement charges for congestion are based on the same nodal prices. This means that congestion charges and transmission rights that serve as congestion hedges are calculated in a consistent manner, using the same grid conditions and assumptions, and both are consistent with the dispatch for the market in which they are settled.

This internal consistency ensures that the FTR settlements do not undermine or distort any of the efficiency properties of the nodal prices and related usage charges. Parties receive efficient price signals from the nodal prices for both short-run operations and long-run investments, and nothing in the FTR settlement rules undermines the efficient signals that each party receives regarding its use of the grid.

FTRs defined in this manner fulfill all of the objectives of Order 2000. They are consistent with FERC’s goals of an efficient dispatch, efficient price signals and efficient allocation of grid use. An RTO implementing the Standard Market Design should therefore offer financial transmission rights that meet these same principles.

- In conjunction with a nodal pricing system, an RTO should offer and provide settlement support for a system of point-to-point financial transmission rights.
- The FTRs are “financial,” in that they entitle the holder to a settlement credit that can offset the congestion-related components of nodal prices or the congestion-based usage charges for a corresponding transaction. The financial characteristics of FTRs are required to facilitate independent rights trading while ensuring that they do not undermine an efficient dispatch or distort efficient price signals:
 - Parties are not required to acquire or hold FTRs as a condition for scheduling or gaining grid access.
 - Parties are not required to match their FTRs to their schedules.
 - Parties are free to participate or not in the ISO’s security-constrained economic dispatch, or to schedule transmission, whether or not they hold FTRs.
- In the RTO settlements, parties receive the market value of the FTRs they hold, whether or not they implement a matching schedule.
- Settlement rules for FTRs should not undermine or distort the efficient price signals provided by the nodal pricing system. FTR settlements should use the same nodal prices that are used to define spot market settlements and usage charges, so that congestion charges and the value of congestion hedges are defined by the same grid conditions and assumptions, and both are consistent with the actual dispatch.⁴⁵

Point-to-Point and Constraint/Flowgate Rights

The FTRs described in the previous sections and offered in PJM and New York are defined from point to point, with their values defined by the nodal prices at each point. The Standard Market Design includes point-to-point FTRs.

Some parties have indicated a desire to manage congestion pricing risks through transmission rights that are defined on a constraint-by-constraint basis. Constraint-specific rights, or “flowgate” rights (FGRs), as they are often called, could also be defined as financial rights.⁴⁶ Whether they could meet each of the foundation principles in this paper, however, remains to be demonstrated. For example, to be compatible with the Standard Market Design, and avoid undermining the efficient price signals achieved by the nodal pricing system, the flowgate rights would need to meet the following requirements:

⁴⁵ If the RTO operates a day-ahead market, the FTRs would be settled in the day-ahead market using day-ahead nodal prices. Day-ahead FTR credits would therefore be consistent with day-ahead nodal prices and usage charges.

⁴⁶ A point-to-point FTR is a complete set of flowgate rights for all of the possible constraints that could be binding between the two grid points defined by the FTR. In theory, therefore, there is a set of flowgate rights that corresponds to each point-to-point FTR. However, this correspondence does not hold as grid conditions change, thus complicating efforts by an RTO to use FTRs and flowgate rights interchangeably. However, nothing in the Standard Design’s inclusion of FTRs would prevent private parties from acquiring FTRs and making their own markets for trading in flowgate rights.

- *FGRs* would be financial rights, and be subject to the same principles regarding their financial attributes that apply to point-to-point FTRs.
- In the RTO settlements, parties would receive the market value of the *FGRs* they hold, whether or not they implemented a matching schedule.
- Settlement rules for *FGRs* should not undermine or distort the efficient price signals provided by the nodal pricing system. To avoid undermining the efficient incentive properties of the nodal prices, settlement rules for *FGRs* should be consistent with those prices, with the value of *FGRs* determined by the same dispatch, grid conditions and assumptions as those used to determine the nodal prices used for imbalance/spot energy settlements and transmission usage charges.

FTR Obligations and Options

The FTRs offered in PJM and New York and similar rights proposed for New England function as “obligations.” The defining feature of an obligation FTR is that the holder of the FTR is not only entitled to receive the difference in nodal prices if the difference is positive, but is obligated to pay (in the RTO settlements) the difference in nodal prices if the difference is negative. This would be the case if the nodal price at the point of receipt defined by the FTR were higher than the nodal price at the point of delivery defined by the FTR.

Obligation FTRs can be used to provide hedges for proposed counter-flows that help to relieve congestion. Counter-flows reduce the marginal cost of any congestion redispatch. In cases where the locational price differences are negative, a party with an obligation FTR would either provide the counterflow (in which case it would receive through the RTO settlements the difference in nodal prices to offset its FTR obligation payments) or pay the marginal redispatch costs (defined as the difference in nodal prices) for the FTR’s locations.

An obligation FTR provides a means to sell congestion management services on a forward basis. Parties willing to undertake the obligation for those locations likely to require payment can therefore be paid to do so. Willing parties are essentially betting that the amount they are paid in advance to take on the obligation will turn out to be greater than the cost of either providing the counter-flows (redispatch) in real time or paying for redispatch. The obligation FTRs that turn out to be “negative” FTRs simply become another product that parties can acquire and trade to support their commercial objectives.

When parties commit either to provide counter-flow/redispatch or to pay for it, these binding commitments allow more transactions to flow in the opposite direction, expanding use of the grid for the entire market. Obligation FTRs thus allow an RTO to maximize grid use, accommodate more transactions and issue more rights.

FTRs can also be defined as “options.” When congestion arises and results in nodal price differences, the option FTR holder is hedged against the congestion/usage charges when the nodal price difference is positive, but the option holder is not obligated to pay the nodal price difference when this difference is negative. Not having this obligation will be valuable to some participants, particular those who do not plan to implement corresponding (counter-flow) schedules. Options may be less important to those prepared to provide the counter-flow or willing to pay for it (pay the marginal redispatch costs). Moreover, if all FTRs are options, no party is obligated either to

provide or to pay for counter-flows through the rights system. Without these commitments, fewer transactions can flow, and the RTO may have to allocate fewer rights if they are all options.

In issuing FTRs, an RTO would use a simultaneous feasibility test, which ensures that the total rights issued can be accommodated on the grid under expected conditions. If the issued rights meet this test, then the RTO would also know that as long as the grid has the capacity assumed under these conditions, it will always collect enough congestion revenues (from nodal prices and usage charges) to fund all of the FTRs, even if the dispatch in each settlement period is different from the dispatch assumed in the simultaneous feasibility test.⁴⁷ That is, congestion revenues collected by the RTO will be adequate to cover FTR payments owed by the RTO.

The linkage between the simultaneous feasibility test and the FTR revenue adequacy condition is an important factor in preserving the quality and value of the FTR hedges, and thus in defining how many FTRs to issue. If the test is not met, revenues may often be inadequate to fund the FTRs, so their hedging value will be diminished. Applying this test is relatively easy if the FTRs are obligations. Until now, however, there has been no workable method for applying the simultaneous feasibility test to FTRs defined as options. This was the primary reason why PJM and New York selected obligations at the time they initiated FTRs. If this problem can be solved, and a method defined to determine how much of each type of right to allocate, there need be no impediment to defining both obligation and option FTRs.⁴⁸ In theory, FTRs could be defined as either obligations or options, and the RTO could issue some of each, depending on the market's preferences.

Even when the simultaneous feasibility test is met, FTR revenue adequacy is not assured when grid conditions change. For example, a line outage that was not accounted for when the FTRs were issued can prevent certain schedules from being implemented, and thus reduce the congestion revenues collected by the RTO. This can result in the RTO having insufficient congestion revenues to fund all of the outstanding FTRs. When revenue inadequacy occurs, the RTO must have some rule to define how to deal with this shortfall.

Several mechanisms are feasible. First, the RTO will collect excess congestion revenue during some settlement periods, and it can use those surpluses to fund revenue shortfalls in other periods. The carryover period might extend a month, six months, or longer. Residual shortfalls could then be handled through a *pro rata* funding of the FTRs. Alternatively, in addition to using surpluses carried over from previous periods, the RTO rules can provide that it will fully fund the FTRs in each settlement period and add any residual deficit or surpluses (after applying surpluses from other periods) to the revenue requirements for the transmission tariff.

Under a third approach, the RTO and its associated grid owner(s) can develop a performance incentive that rewards the grid owner(s) for proper grid maintenance that minimizes grid outages, while holding the owner(s) responsible for FTR revenue shortfalls attributed to grid outages.

⁴⁷Scott M. Harvey, Susan L. Pope, and William Hogan, "Transmission Capacity Reservations and Transmission Congestion Contracts," June 6, 1996.

⁴⁸ In approving the redesign of New England's congestion management markets to implement nodal pricing, FERC directed the ISO-NE to offer both obligation and option FTRs. Work to test a possible approach to the simultaneous feasibility test for FTR options is underway but not yet completed.

FTR funding is thus assured, and the mechanism can encourage efficient grid maintenance through performance incentives.⁴⁹

Financial Transmission Rights and Market-Driven Transmission Investments

An important attribute of FTRs is that they are tradable property rights that reflect the market value of using the grid. They can be acquired in advance through RTO allocations and auctions, and once acquired, they can be traded in forward secondary markets. The prices at which parties acquire and trade FTRs in advance indicate the market's forward valuation of grid usage. Parties will tend to acquire and trade FTRs at prices they think represent the expected value of the stream of payments the holder will receive over the FTR period. The prices will tend to reflect the expected value that the market places on avoiding congestion charges. FTRs thus provide a measure of the market value of upgrading the grid to relieve the same congestion.

Because forward FTR prices can signal the market's economic value of grid improvements, the ability to acquire FTRs can be a mechanism to support market-driven investments in grid upgrades. The award of FTRs thus provides an opportunity for investors to receive the market value of their grid upgrades. An important rule in the Standard Market Design would therefore award to those who fund grid upgrades the incremental FTRs that become possible as a result of those upgrades.⁵⁰

- An RTO should award to those who invest in grid upgrades the incremental FTRs that the upgrades make possible under the simultaneous feasibility test. The length of the FTRs should be consistent with the longevity of the investment, or in any event, long enough to allow a reasonable period for recovery of the cost of the investment and profit expectations.
- An RTO should have a defined method for determining the incremental FTRs made possible by a proposed upgrade.

These principles will allow and encourage market-driven investments in transmission. Many such investments may be possible, and over time they may become an important mechanism for funding transmission expansions.⁵¹ However, it can be expected that market failures will prevent some economically justified transmission investments from being pursued in response to market price incentives. For example, economies of scale may warrant investments greater than can be supported by the market-driven investment at any given moment, and free-rider problems will arise where upgrades provide benefits to many parties, making the formation of voluntary investment coalitions somewhat difficult.

Some form of regulatory backstop is therefore needed for an indefinite period to ensure that appropriate grid investments occur in the face of market failures. The investments would be subject to recovery under traditional transmission tariffs. However, regulated investments should

⁴⁹ Commonwealth Edison proposed such an incentive mechanism in its proposal for an Independent Transmission Company. See, Attachment 1 to "Petition for Declaratory Order," filed December 13, 1999.

⁵⁰ FTRs would be awarded to any entity that funded the investments, whether the original grid owners, a newly formed transco or independent transmission company, a merchant developer or a merely a group of independent investors.

⁵¹ This means that the fixed costs of the grid will very gradually move from a regulated rate structure to a market-based pricing structure.

not be undertaken unless they can be economically justified by benefits defined in relation to expected changes in nodal prices and related usage charges, and only when the regulator can determine that one or more market failures make it unlikely the market will undertake the necessary upgrade investments.

Additional Features Compatible with the Standard Design

This paper describes the essential foundation principles for a Standard Market Design. These are the minimum requirements. However, both existing and proposed markets include additional features that have been requested by market participants and proven helpful in making the markets more workable. Without attempting to describe all desirable features, the following common features can be provided by the RTO and be consistent with the Standard Design.

- *Day-Ahead and/or Hour-Ahead Markets.* In PJM, New York and New England, participants have generally supported having the ISO coordinate bid-based short-run forward markets. Participation in these forward markets is more or less voluntary. The markets are based on participant offers and bids and arranged around an ISO-coordinated security-constrained economic dispatch. The ISO uses the price offers and bids to clear each market and define market-clearing prices for energy bought and sold in each market. Prices are defined using locational marginal pricing, so the pricing reflects the marginal cost of any redispatch needed to accommodate the forward market schedules. Parties that commit to buy and sell energy in the forward market are held financially responsible for these purchases and sales and settled in that market at the forward market-clearing prices. In the real-time market settlements, deviations from the quantities committed to in forward market are then settled at the real-time market prices.
- *Unit Commitment Service.* In PJM, New York and New England (proposed), the ISO offers a voluntary unit commitment service, based on three-part bids, which indicate each plant's incremental energy costs, start-up costs and minimum generation costs. Generators are allowed to self-schedule their own units, but they may also allow the ISO to determine the most economic unit commitment for their plants. Participating generators are guaranteed recovery of their start-up and minimum generation costs in the event they fail to recover these costs from the prices received in the ISO-coordinated markets.

Mitigation of Market Power

The potential for market power will need to be addressed in each market region. This is a complex topic well beyond the scope of this paper. In general, the exercise of market power is more likely and can be more profitable under a defective market design; hence close adherence to the principles in the Standard Market Design is an essential strategy for minimizing the risks of market power. While it may be easier to deal with market power issues prior to market commencement, it is likely that grid constraints will still leave pockets in which strategic bidding or withholding to raise prices above efficient levels can be successful if not mitigated. These concerns can be addressed through mandatory availability requirements, forward contracts for "reliability-must-run" units, and selective bid caps on generators at locations where grid constraints allow the exercise of market power. The availability requirements ensure that units essential for reliability are offered to the RTO through bids into its real-time markets. The bid caps serve to limit the prices that can be offered by those in strategic grid locations where,

because of grid constraints, little or no effective competition exists. The Standard Market Design simplifies market power mitigation by allowing mitigation rules that would be consistent with a competitive market.

Related Issues

The principles described in this paper provide much of the essential foundation for a workable, competitive electricity market. However, there are many other design details that will affect how well a market will function. An especially important topic not discussed in this paper concerns the method by which market prices are defined and possibly limited during periods of supply shortages, how demand bids can be used to define prices, and how prices are set during real shortage conditions in the absence of demand-side responses. A related topic concerns the merits of capacity requirements and the possible use of capacity markets or other mechanisms to help assure supply adequacy. These are extremely complex topics for which one or more separate papers would be appropriate. Additional papers would discuss such topics as state regulatory actions to support markets, the complementary role of retail pricing and the conditions for enhancing demand-side responses to market prices. These issues can be examined intelligently within a framework in which the core principles described in this paper are accepted as the foundation. Without the foundation provided by the Standard Market Design, discussion of these and other obviously important topics would have much less value.

Conclusion

FERC's goals for efficient competitive markets would be substantially furthered if RTOs were strongly encouraged to adhere to the foundation principles in the Standard Market Design. The principles define RTO market support functions and pricing methods that have a proven record of success in the Eastern United States and elsewhere. The Standard Design is based on a bid-based, security-constrained economic dispatch, coordinated by the RTO, with the market prices used for settlement defined by locational marginal costs. Financial Transmission Rights that hedge locational price differences and can be traded and acquired in advance to provide price certainty complete the basic design and provide a basis for market-driven grid investments and trading of risk management instruments. Within this Standard Design framework, participants can freely engage in bilateral trading, self scheduling and other decentralized trading or rely on coordinated spot trading settled at market-clearing prices defined by the RTO. Trading can then be simplified and made more liquid by defining financial trading hubs or other nodal pricing aggregations for settlements.

Experience with markets based on these principles is sufficient to establish them as the benchmark for RTO functions and "best practice" market designs. Further innovation and experiments with design principles can be considered, but proponents should bear the burden of showing that untried approaches are likely to work, pose acceptable risks and are more likely to achieve the goals of an efficient competitive market than the Standard Design.