



Andrew L. Ott
Vice President, Markets

955 Jefferson Avenue
Norristown, PA 19403
Office: 610-666-4267
Fax: 610-666-4281
ott@pjm.com

How RTOs Establish Spot Market Prices (And How This Helps To Keep The Lights On)

This paper explains in basic terms how wholesale spot energy market prices are established by Regional Transmission Organizations (RTOs) such as PJM and are an integral and necessary part of the mechanisms RTOs in the United States use to ensure the reliable operation of the electric transmission system. The paper also discusses the relationship of “market clearing prices” used in “uniform price auctions” as well as the need for the use of “locational market prices”. It is written for individuals who have some knowledge of RTOs and independent system operators (ISOs).

This paper which was commissioned by PJM Interconnection and authored by Mr. John Chandley, a principal in the electricity market design group of LECG, is part of the effort by PJM to continue to explain how PJM’s electricity market - which includes the day ahead and real time energy market – helps PJM to keep the lights on.

Andy Ott
Vice President – Markets
PJM Interconnection
September, 2007

How RTOs Set Spot Market Prices (And How It Helps Keep the Lights On)¹

John Chandley, LECG, LLC.
September 2007

Introduction

More than half the nation's electricity grid is now operated by large Regional Transmission Organizations.² As shown in the Appendix, Figure 1, these RTOs ensure reliable electricity service for all of the Eastern US except for the South; for most of the Midwest, Texas and small parts of the Southwest; and for most of California. In addition to ensuring reliable operations, RTOs also operate electricity "spot markets" in which generators sell and utilities and customers buy "spot energy," energy that is bought and sold at or very close to the time in which it is produced and consumed.³ However, the spot markets are not simply for commercial trading; they are also an integral part of the mechanisms every RTO uses to ensure reliable operations.

¹ This paper was commissioned by PJM Interconnection. The views expressed are my own.

² A Regional Transmission Organization – RTO – is essentially an organization that, among other things, *operates a regional power pool*. Power pools have been around almost since the beginning of electricity power systems – PJM began in 1927 -- and are common throughout the industry, in all countries. Using the transmission system to move power, virtually every electric utility "pools" its generators and coordinates their operations to serve the collective demands of the utility's customers, rather than having specific generators serve only specific customers. Utility pools typically serve only their customers; RTO power pools tend to consolidate the dispatch and transmission operations of multiple utilities across a larger region, but the "pooling" concept is essentially the same: coordinated dispatch of all generators in the "pool" to serve all the demands (loads) in the aggregate system. Because RTOs operate transmission systems that cross state boundaries and thus move power in interstate commerce, their operations are subject to oversight by the Federal Energy Regulatory Commission, while most of the retail rates for the individual utilities within the RTO pool are regulated exclusively by each state.

³ RTOs typically operate both a day-ahead spot market and a real-time spot market. Parties can buy and sell energy in the day-ahead market, locking in day-ahead prices for the energy they expect to produce or consume the next day; or they can buy and sell energy on an hourly or even shorter basis in the real-time market, with prices set in "real time" – that is, at the moment the energy is physically generated and consumed. In addition to using these RTO-coordinated spot markets, parties in RTO regions can "self-schedule" their own power plants to serve their own customer loads, and they can engage in bilateral contracts to buy and sell electricity for periods ranging from hours before real time to weeks, months or even years. RTOs accommodate these self schedules and contractual arrangements as they arrange real-time operations, while using the spot markets to allow parties to cover any imbalances or differences between their announced self schedules and forward contracts and their actual production or consumption in real time. For this reason, the RTO real-time spot markets are sometimes called "balancing markets."

Though now a common and accepted feature of every RTO, the spot markets are one of the most misunderstood aspects of RTO operations. Some misconceptions involve how the RTOs set the “market-clearing” prices in the hourly power auctions that RTOs use in their spot markets. These auctions operate every day and rely on offers to sell energy from suppliers/generators and bids to purchase from buyers, such as utilities, retailers and ultimate consumers. Each RTO conducts an auction for each hour of the operating day, in both the day-ahead and real-time periods. For example there is a day-ahead market auction for the hour beginning 7:00 a.m. tomorrow, and a real-time market auction for that same hour conducted on the operating day. The same is true for every hour of the day.

RTO spot markets function at the “wholesale” level; utilities and competitive retailers who purchase energy from these wholesale markets then resell it to final consumers at retail rates set by state regulators. However, in some states, especially those with advanced demand-side management programs, retail rates for larger customers may be based on a more or less direct pass through of the hourly wholesale spot prices (or close approximations), blurring this distinction between wholesale and retail. These rate designs are still set by state regulators. Moreover, over time, retail rates for all customers tend to reflect average wholesale costs; so an RTO’s spot prices are related to eventual retail rates, even if there is no direct and immediate pass through of hourly wholesale spot prices.

Using auctions both to acquire supplies through competition and to set market-clearing prices is fairly common. Under a commonly used “uniform price auction” method, the spot price for energy for the hour (or shorter dispatch interval) is set by the intersection of supply offers and demand bids, and all purchases and sales occurring at the same time and location are settled at that same price – the “uniform” price. That uniform price, however, can change from one time interval to the next and from one location to another.

A common misconception is that a uniform price auction results in consumers paying higher prices than they should, even though (as explained below) there is no theoretical, logical or empirical basis for this concern. A related misconception holds that this approach always allows the highest cost resource⁴ on the system to set the clearing price, while another is that the resulting prices provide unjust enrichment to low-cost plants that would otherwise consistently be willing to sell their output at much lower prices. These and related notions are simply false – they are misconceptions about how the spot markets work, how an RTO sets spot market prices, how market participants behave depending on the incentives in the pricing rules, and the critical role spot market prices play in the overall scheme of keeping the lights on. This paper will address each of these ideas.

The truth is that after initially and unsuccessfully trying other approaches to setting spot market prices, all US RTOs now agree on the uniform price approach and all use a common, successfully tested method for setting spot market prices at each location on the grid. This common approach – called Locational Marginal Pricing, or LMP -- ensures that electricity consumers are reliably served at the lowest-cost combination of power plants and demand-side resources available to the RTO. Moreover, the RTO’s spot market prices provide price signals at each location that consistently encourage generators and price-responsive demand-side resources to follow the RTO’s dispatch instructions.⁵ In sum, correctly set spot prices encourage all parties to do exactly what the RTO needs them to do to keep the lights on at every moment.

⁴ The term “resource” in this paper includes generators who provide energy to the grid, but it also includes “demand-side” resources, which involve efforts by ultimate consumers (or those who aggregate such efforts) to curtail demand in response to requests by the RTO system operators. Balancing supply and demand can thus be accomplished by either generation resources or demand-side resources.

⁵ The “dispatch” is the process every electricity system operator uses to keep the lights on. Every system operator, whether a utility or an RTO, coordinates a dispatch to determine which resources to call up at each moment to match supply and demand and maintain reliable operations. After selecting the lowest cost set of providers to participate in the dispatch, a system operator sends “dispatch instructions” to each provider telling them when and how much energy to produce (or for demand-side resource providers, how much to reduce consumption). As long as participants follow dispatch instructions, the system stays in balance, frequency and voltage requirements are met, and the lights stay on.

Far from raising costs and unfairly rewarding generators, RTO spot market pricing actually does the opposite: by accurately pricing the RTO's dispatch, such pricing helps ensure reliability in both the short and long run. In the short run, RTO spot prices encourage generation and demand-side resources to participate in the dispatch and follow dispatch instructions. In the long run, the spot market prices reflect, over time, the cost of maintaining reliability with the lowest cost set of resources that can meet the total system's reliability needs, while pricing the energy (and operating reserves) at a level consistent with long-run investment requirements.

RTOs have tried alternative pricing approaches without success. When PJM (1997), California (1998), ERCOT and New England (2000-01) began market operations, each tried pricing methods different from the method commonly accepted today. Some (notably California in 2000-2001) tried pricing approaches based on pay-as-bid schemes.⁶ But most pay-as-bid schemes have been abandoned in favor of the uniform price auction approach. Some RTOs used approaches that ignored locational differences in marginal costs, even though it costs more to serve demand in some locations than it does in others; such differences are created by transmission limitations into many areas. California and ERCOT tried "zonal pricing" compromised to recognize a few but not all of these locational differences;⁷ both RTOs gave up on zonal pricing, because it became unworkable. Each of these efforts caused total dispatch costs to rise and/or made system reliability more difficult to maintain. All RTOs now agree that using uniform price

⁶ In a pay-as-bid auction, each bidder selected for dispatch is paid the price it bid into the auction, rather than having all successful bidders paid the clearing price regardless of their bids. Comparisons between these approaches are only valid when one recognizes that bidders bid differently in the two schemes.

⁷ In a "zonal pricing" scheme, the RTO attempts to define large subregions or "zones" within the RTO footprint. Congestion on lines connecting the zones is supposedly significant, whereas congestion within each zone is thought to be insignificant or non-existent. Prices between the zones can therefore differ, but prices within any given zone are, by RTO rule, forced to be the same. The problem is that these simplifying assumptions seldom hold; actual grid congestion is not easily predictable, congestion arises unexpectedly inside a zone, and using the same price within a zone that has congestion tends to exacerbate the congestion, while encouraging generators to "game" their bids or induce congestion to maximize profits. The early California, Texas, UK and Ontario markets all suffered from these problems with zonal or one-zone pricing. California and Texas are now implementing LMP; Ontario imposes a uniform price for the entire Province and buries the problems with imposed administrative prices, and the UK essentially gave up on having an open spot market because its non-LMP pricing system led to persistent "gaming" of generator bids.

auctions to set spot market energy prices at the marginal cost of the dispatch at each location works to support reliability and allows least-cost operations. The basic approach of locational marginal pricing is explained further below.

Despite starting with different approaches, every RTO eventually chose the common spot market pricing approach; each came to that decision voluntarily, without federal mandates and sometimes despite federal indifference, delays and conflicting guidance. But to see why they did so, it is first necessary to explore why RTOs even operate spot markets and how the way they set spot market prices matters to the system operators whose job is to ensure reliability.

Why Do All RTOs Operate Electricity Spot Markets?

An RTO's core function is to maintain reliability across the system; it does this by coordinating the dispatch of generators and demand-side providers. By sending dispatch instructions to selected generators and demand-side resource providers, the RTO precisely balances supply and demand, bringing more supply on line as demand rises (or adjusting consumption by demand side providers) and taking supplies off as demand declines. These adjustments in the dispatch occur every five minutes, all day long, every day – 24/7. Even more refined adjustments (called “regulation” or “automatic generation control” – AGC) occur during each five-minute interval to maintain an extremely precise balance between supply and demand ever second. A dispatch that keeps the lights on by maintaining constant balance, while keeping electricity flows across each transmission element within safe operating limits, and does this at the lowest cost, is called a “security-constrained, economic dispatch.” *Every utility and RTO system operator in the world utilizes a security-constrained, economic dispatch.*

RTOs operate large, regional power pools that combine the dispatch across several/many utilities in a larger region. See Appendix, Figure 2. And because they operate power pools across multiple utility systems, they must make their combined

dispatch open and available to all generators in the region regardless of ownership. RTOs thus operate “open” power pools.⁸

An RTO that operates an “open pool” must also operate a spot market. Operating a real-time spot market is the unavoidable and natural consequence of the open power pool dispatch that every RTO must use to maintain reliability. The spot market also arises partly from an RTO’s independent status.⁹ The RTO is an independent organization; it does not own power plants nor is it a load-serving utility responsible for consumer demand. Except in emergency situations, it cannot routinely order generators from various companies what to do. Instead, RTOs use economic incentives – spot market prices derived from the dispatch -- to induce parties to comply with dispatch instructions.¹⁰

To arrange a least-cost dispatch, the RTO must compare the costs (as reflected in their offers) of dispatching different generators that may be offering their energy for dispatch. An RTO accepts price-quantity offers (in \$/MWh)¹¹ to sell energy from

⁸ An “open power pool” is one in which any generator can participate in the power pool’s security-constrained economic dispatch. In contrast, many utility power pools are closed; they do not automatically allow non-utility power plants to participate in the dispatch and may exclude third parties, even though those excluded may be able to produce power more cheaply than the most costly utility plants. An RTO operates an “open” power pool by allowing any generator, regardless of ownership or utility affiliation, to submit offers to participate in the dispatch. The RTO selects the least-cost mix of generation to meet the dispatch requirements, and it pays all participating generators under the same non-discriminatory pricing rules. This is a major difference between utility-controlled systems and pools operated by RTOs.

⁹ RTOs are quasi-public organizations. They are “independent” in the sense of having no financial affiliation with any market participant; they don’t own generation and are not owned or controlled by generation companies. US RTOs are also independent from transmission owning utilities; in some countries, however, there are transmission companies that perform RTO-type dispatch functions while remaining independent of generation ownership.

¹⁰ As grid/system operators responsible for dispatch/reliability over wide regions, RTOs also have the authority (backed by severe penalties for non-compliance) to order parties to comply with dispatch instructions during extreme emergency conditions, when normal economic incentives may be insufficient to avoid system collapse. These conditions are rare. During normal operations, the RTO relies on economic incentives provided by spot market prices to induce appropriate dispatch behavior.

¹¹ E.g., each generator may submit a series of offer prices or a supply curve showing how much energy it would produce at increasing price levels if selected for dispatch.

generators and price-quantity bids to purchase energy from loads.¹² Using these offers and bids, the RTO selects the lowest-cost combination to serve the system load at or below the prices loads are willing to pay. See Appendix, Figures 3 and 4. Once generators produce energy in response to dispatch instructions, the RTO must pay generators for their energy output, and it must charge loads for the energy they take. It follows that a system to price energy as it is produced (sold) and consumed (bought) – a spot market – is the natural consequence of an independent RTO operating a regional dispatch. See Figure 5.

In the absence of extensive mandatory rules that participants willingly accept,¹³ it is not practical for an RTO that does not own generators to operate a power pool and to dispatch enough generation to reliably meet demand unless the RTO also operates a spot market that determines the correct price to pay generators and pays them that price. And the “correct” prices are also important to bidding loads, because their bids indicate none of them would be willing to pay more for energy than the price bids they submit. Price-responsive demand would not participate nearly as much unless the price savings rewarded them for the value of their demand reductions. So an unavoidable corollary to the necessity of operating a spot market is that the RTO must price energy from the spot market in a way that ensures generators come on line and follow dispatch signals, and ensures loads will pay enough to cover this cost, while paying no more than they were willing to pay, as indicated by their bids.

¹² The term “load” is used here to mean an entity that either consumes energy from the grid or purchases energy to serve its customers. In this context “load” is associated with the demand for energy.

¹³ Prior to today’s spot markets, early power pools, such as PJM, were able to secure compliance with dispatch instructions because the limited number of member utilities comprising the pool voluntarily agreed to follow the PJM-pool dispatch instructions. In exchange, the members agreed to share the savings that pooled dispatch created compared to what separate utility dispatches would have cost. These split-savings approaches were rudimentary spot markets that tolerated a significant degree of cost-shifting between members; their pricing rules were extremely complicated and required voluntary compliance by pool members. Opening the pool dispatch to many utilities and dozens of independent generation owners – as we find today in PJM and Midwest ISO -- made such voluntary systems unworkable, and the cost-shifting became unacceptable. That is why no RTO still uses a “split-the-savings” approach, even though early power pools started with that mechanism.

Another way of saying this is that spot market prices must provide the correct incentives, so that all players, generators and loads, do what the RTO needs them to do to keep the lights on. Spot market pricing must support reliability, and if it doesn't, dispatchable generators and loads might cease to follow dispatch instructions to avoid losing money. When that happens, reliability can fail and the lights can go out. *In RTOs, spot market pricing and maintaining reliability are inextricably linked.*

The fact that RTOs must operate spot markets for buying and selling electricity in conjunction with the real-time dispatch was not always understood. A decade ago, when basic RTO rules were being debated, some argued that RTOs should focus exclusively on reliability, or “operating the grid,” but they should not operate a market. The argument was that the workings of the “market” should be left to market participants. This concept was promoted by Enron and others, first in California and then elsewhere. These parties explicitly sought to enhance profit-making arbitrage opportunities for marketers, but the concept proved to be completely unworkable, as well as susceptible to widespread abuse. Moreover, the notion that an RTO should not try to minimize dispatch costs, but instead deliberately create an inefficient dispatch so as to create profit-making opportunities for marketers, is inherently inefficient and anti-consumer. Every US RTO now rejects such notions.

No RTO can easily keep the lights on and support an efficient market without also operating a properly functioning spot market, with prices correctly set to support a reliable dispatch. The real-time dispatch that keeps the lights on, and the real-time spot market that manages the financing of the dispatch, are two sides – operational and financial – of the same integrated function. But this integrated reliability function only works well if the RTO sets prices correctly – consistent with the offers/bids, consistent with the dispatch, and consistent with the actual grid conditions that affect the dispatch.

There are examples in other countries (and early US efforts) where spot markets do not work well. In those cases, the system operator must still maintain reliability, and so it must counteract the perverse incentives of spot prices that are inconsistent with

reliable dispatch. In such cases, a system operator can maintain reliability only by imposing a set of mandatory rules enforced by mandatory penalties for non-compliance, mechanisms that tend to increase total costs of a reliable dispatch. These systems are inherently inefficient. The Brazilian and British systems are examples of this. In other countries, such as New Zealand, and to a lesser extent Australia and the Canadian Province of Ontario, better designed spot markets tend to support reliable dispatch with better spot pricing rules and fewer mandatory rules. There are variations in all these systems, but the better markets are those that understand the importance of open spot markets and the need for good spot pricing rules that encourage generators to follow dispatch instructions.

How Do RTOs Set Spot Energy Prices?

RTOs must set spot market prices in a way that supports both reliable dispatch *and* least-cost dispatch, with “least-cost” defined as the lowest as-bid cost for the set of generators and demand-side providers chosen, as determined from the price offers and bids the RTO receives for the dispatch. More broadly, the RTO must set spot prices in a way that meets the following conditions:

- 1. To induce generators to contribute to the dispatch, RTO spot prices must reflect the **market value** of the energy each generator produces at the time and place it is produced.*

Every generator in an RTO system has the option to sell its power exclusively through bilateral contracts with utilities, wholesalers, competitive retailers and so on – or even to neighboring power pools.¹⁴ Yet an RTO needs to convince enough of these generators to also participate in the RTO’s dispatch, so that the RTO can use the dispatch to maintain system-wide energy balance, voltage support and stable frequency. When a generator

¹⁴ For example, a generator located in PJM can sell its power in the Midwest ISO spot markets, because the two RTOs are closely interconnected and each RTO’s dispatch is open to purchases and sales from any seller or buyer. Even before RTOs and regional power pools, interconnected utilities engaged in “off-system” sales/purchases with entities outside the utility’s service territory, that would benefit both buyer and seller. RTOs now increase and facilitate such beneficial trading by accommodating the inter-utility transmission use and using the RTO dispatch to cover any imbalances or need for redispatch, while pricing the dispatch services through the spot market.

participates in an RTO's dispatch, it is selling its energy output through the dispatch at the spot market price. But no generator would voluntarily sell its energy output in the RTO spot market/dispatch unless the spot price reflected the market value – the price that the generator could receive if it sold its power into these other markets. And because the market value of energy depends on marginal costs when and where it is produced, the RTO spot prices must reflect changes in market value over the day and at different locations. This means that spot electricity prices are inherently volatile: they change from hour to hour and from location to location.¹⁵

2. *To support reliable dispatch, the spot market price must be consistent with the generators' offers (and the demand-response providers' bids).*

Because generators and loads can sell into or buy out of other markets, no party that submits an offer or bid to the RTO should regret following the RTO's dispatch instructions and being paid/charged at the resulting spot prices. Thus, if a generator submits an offer and the RTO selects that offer (because it is part of the lowest-cost set of supplies to meet demand) and the RTO instructs the generator to produce energy, the generator will receive a price that is equal to or greater than its offer price. The generator will not lose money (or receive at least its offer price) as a result of following the RTO's dispatch instructions and accepting the resulting spot price, because the spot price will be high enough to at least cover the generator's price offer, which is likely to reflect its operating costs for producing the energy requested by the RTO. Similarly, no load should regret purchasing from the spot market, because the price it pays will be no greater than the bid it submitted. The principle might be called a "no regrets" rule.

3. *To achieve least cost dispatch, the spot market price should encourage parties to offer/bid in ways that support least-cost dispatch.*

¹⁵ There is nothing unusual about the volatility of spot electricity prices. The marginal cost of generation, and hence the market value, is higher during periods of high demand (peak hours each day) than at non-peak demand hours, because more expensive generators must be brought on line as demand increases. Market value is typically higher in transmission constrained areas than it is in areas with unconstrained transmission. Spot prices, which reflect marginal costs, therefore tend to be volatile because different plants with different marginal operating costs are on the margin – and setting the spot price -- at different times and different locations.

Because generators know they will receive spot market prices set by the “no-regrets” rule above, this principle also tends to encourage generators to submit offers at or near their marginal operating costs, or their opportunity costs – the profits they forego by not selling into another market. In other words, the rule encourages generators to submit offers at the lowest price at which they would be willing to sell their power. The result is a “least-cost” dispatch.

This is the point where debates about “uniform-price” auctions versus “pay-as-bid” schemes arise, because pay-as-bid supporters neglect a fundamental economic fact: *Parties who are free to decide what to offer/bid will change their price offers/bids depending on the incentives provided by the pricing mechanism.* Uniform price auctions and pay-as-bid auctions produce very different incentives, and thus encourage very different bidding behavior.

- In general, if the spot pricing rule says the RTO will pay generators what they offer (“pay-as-bid”), then generators will submit offers reflecting what they want to receive – i.e., their expectations of the market clearing price. In other words, they will bid what they think the power is worth, not what it costs to produce. They will not submit offers at only their marginal operating costs.
- If, on the other hand, the spot pricing rule says the RTO will pay the generators the clearing price, then generators will tend to submit offers that reflect their marginal operating costs or their opportunity costs. In other words, they will tend to bid their costs, in order to assure they are selected for dispatch, while knowing that the clearing price in a uniform price auction will reflect what the power is worth.

These very different and predictable effects, for which there is much confirming experience, mean that in most cases *generator offers will be lower* in a clearing price system than they would be in system that uses a “pay as bid” rule.¹⁶ But *spot prices will*

¹⁶ There are numerous economic studies of the different incentives provided by clearing price rules versus pay-as-bid rules. In the electricity field, the most useful are (1) *Pricing in the California Power Exchange Electricity Market: Should California Switch from Uniform Pricing to Pay-as-Bid Pricing?*, a study by a “blue-ribbon panel” of experts for the California Power Exchange during the 2001 electricity crisis, when the state was considering changing to a pay- as-bid system; (2) a paper by Peter Cramton and Steven Stoft, *Uniform-Price Auctions in Electricity Markets*, March 18, 2006; and (3) Rebuttal Testimony of William

not be any higher in the clearing price system and may even be lower, because there is no incentive to guess the clearing price – and hence guess wrong -- when submitting offers under the clearing price approach. **The common assumption that a clearing price system that uses a uniform price auction results in higher prices than pay-as-bid approaches is not supported by either economic theory or actual experiences in electricity spot markets where pay-as-bid schemes have been tried.**

RTOs need generators to submit offers reflecting their marginal operating costs, because that is what allows the RTO to arrange a least-cost dispatch. If generators instead submitted offers reflecting their expectations of the market price (as pay-as-bid systems tend to encourage), rather than their operating costs, then many offers would tend to cluster around the generators' common expectation of that market price, which would make it difficult for the RTO to sort out the lowest-cost providers and arrange a least-cost dispatch.

4. To ensure reliable dispatch in the face of transmission congestion, the spot market prices must reflect locational differences caused by the congestion.

Transmission congestion is a common condition on all modern electrical grids. It occurs when it would be cheaper to move more power across a limited transmission line or element than it can accommodate and still meet safe operating standards. Persistent violations could undermine reliable operations and in the extreme, cause transmission outages, widespread damage and/or blackouts. To prevent safety violations from happening, the dispatch operator – the RTO – must frequently adjust the dispatch (“redispatch”) to change the pattern of flows that would otherwise occur. These adjustments deviate from the lowest-cost unconstrained dispatch just enough to avoid overloading the constrained lines or equipment. Redispatch is a normal part of any dispatch – it defines “security-constrained” dispatch. But it changes the mix of

Hogan, filed July 6, 2005 for the Illinois Commerce Commission consideration of default supply auctions, Docket No. 05-0159. As recommended by these papers, all RTOs follow the basic “clearing price” rule.

generators that are dispatched in any given 5-minute dispatch interval, and thus changes the mix of generators whose offers (marginal costs) set the spot market price.¹⁷

Moreover, transmission congestion prevents lower-cost generators from moving power to higher-cost regions, producing a change in the clearing prices at each location. Whenever this happens, *the widely held assumption that the clearing price is set by the highest cost generator being dispatched is no longer true*. If there were no congestion and no losses on the grid, the highest offer would be the marginal cost and would set the spot market price everywhere. But there are always losses, and the need for congestion redispatch is frequent and pervasive. When these conditions require redispatch, spot market clearing prices will differ across the system. See Figure 5.

- At some locations, the clearing price can be *higher* than any generator offer;
- At other locations, the clearing price can be *lower* than any offer;
- At still other locations, prices may *vary in between* high and low offers.¹⁸
- In some cases, the clearing price at a location can even be *zero or negative* (implying we should pay customers to consume energy at those locations because doing so will reduce total costs by making expensive redispatch less necessary).
A few negative prices can happen on any day, somewhere on the system.¹⁹

5. To ensure the lowest total cost of operating the system reliably, spot market energy prices must reflect the tradeoffs between producing energy and providing operating reserves.

¹⁷ Spot prices are typically defined by the offers of the generating plants that are on the margin of the dispatch. Being “on the margin” means that during a dispatch interval (each five minutes) a plant’s output is being adjusted – up or down – in accordance with RTO dispatch instructions, in order to serve the next increment of demand/load while honoring all transmission limits. More than one plant can be on the margin when the system experiences congestion.

¹⁸ This can occur because more than one generating plant is on the margin and affecting the marginal cost of serving load at that location. This occurs because the RTO must “redispatch” lower cost plants down and higher cost plants up. It is the interaction between all of the plants on the margin during redispatch that determines the LMP at a location. If only one plant is on that margin, that plant’s offer will set the price, but because of congestion and redispatch, multiple plants may be on the margin simultaneously.

¹⁹ See <http://www.pjm.com/markets/energy-market/day-ahead.html> to observe hourly changes in LMP.

Every electricity system, without exception, needs operating reserves. A generator can produce power at a price, or it can provide operating reserves – meaning it is standing by ready to produce power if the RTO needs it when another plant or transmission line experiences a sudden and unexpected outage. For each portion of its capacity, a plant can either produce energy or provide operating reserves, but not both simultaneously for the same portion of capacity. For example, a plant with a capacity of 100 megawatts (MW) can use that capacity to produce energy or use it to provide reserves. It can allocate 50 MW of its capacity to produce energy and 50 MW to stand by for reserves, or any other combination; but whatever capacity it uses for energy is, by definition, not available to stand by as operating reserves, and *vice versa*.

Although energy and operating reserves are therefore different products, this tradeoff creates a connection between the spot market price of energy and the market price of operating reserves. For example, the operating reserve clearing price for a plant held back for operating reserves must reflect the opportunity cost that plant incurs – the profits it cannot earn -- because it is not operating and thus not receiving the clearing prices from the energy spot market. The clearing price for energy and operating reserves are thus related and can affect each other. This is another reason why the simple notion that spot prices are set by the most expensive plants dispatched is often not true, because the spot price of energy can be affected by what is happening with operating reserves.

As a general rule, if the RTO does not have enough capacity committed to the dispatch to meet its operating reserve requirements, this “shortage” of operating reserves will tend to raise the spot market price for energy. Consumers may be fully served, with no blackouts, because the RTO has enough capacity to provide all the energy demanded by consumers, and the RTO dispatch will automatically meet that demand for energy. But the market value – and hence the price -- of that energy will rise when the RTO has an operating reserve shortage, because the production of energy is coming at the added cost of lower operating reserves than desired, a condition that is diminishing the level of

reliability. The added cost of short-changing operating reserves raises the price of energy. See Appendix, Figure 8 for an illustration of the concept.²⁰

An RTO's spot pricing rules must satisfy all of these conditions in order to achieve reliable operations at the lowest overall cost for operating the system. There is only one set of pricing rules that consistently satisfies these conditions, achieving least cost and supporting reliability. That system is called, "locational marginal pricing," or LMP. LMP is now being used (or in the process of being implemented) by all RTOs in the United States (and several other countries). The following RTOs have endorsed/implemented LMP:

- PJM – operational since 1998.
- New York Independent System Operator – since 1999
- ISO-New England – since 2003
- Midwest ISO – since 2005
- Southwest Power Pool – since 2007 though incomplete
- California ISO – approved; planned for 2008
- ERCOT (ISO for most of Texas) – approved; planned for 2008-09

Figure 6 in the Appendix illustrates the components of an LMP-based system, based on offers and bids to a security-constrained dispatch, spot market prices derived from the dispatch and the offers/bids, and ability to use the open dispatch to provide balancing for schedules, buying and selling of spot energy, and transmission usage charges based on the marginal costs of redispatch..

How and Why LMP Spot Market Pricing Works

There are several attributes that define a system that uses locational marginal pricing (LMP). These attributes ensure that the system meets the conditions of supporting reliability at the lowest as-bid costs.

²⁰ The "costs" of reserve shortages can be depicted by a downward sloping curve for operating reserves. If the RTO had adequate operating reserves, there would be no added costs for reduced reliability, and hence energy prices would not rise. See, Figure 8.

1. LMP reflects the marginal cost of the dispatch.

The LMP for each location reflects the marginal cost of serving an increment (1 more MW) of load at that location. To calculate the LMP at location A (e.g., a generator's bus/interconnection to the grid, or a transmission substation serving a town), the LMP processor (an automated computer program fed by data from the actual dispatch) looks at the actual dispatch and the offer/bid prices during that dispatch interval. It then calculates how much it would cost to serve one more megawatt of demand at, e.g., location A, given the actual dispatch, the offers/bids, and the transmission constraints.²¹ The incremental change in the dispatch costs for serving that increment of demand is the locational marginal price (LMP) at that location.

An RTO's LMP processor quickly and automatically calculates the LMP for each location on the grid, in a process that takes a couple of minutes after each five-minute dispatch interval. For the real-time spot market, this allows the RTO to post the LMPs on its web site every five minutes following each dispatch interval. In effect, anyone can then see the real-time spot market prices as they change every five minutes at numerous locations. See, e.g., this website: <http://www.pjm.com/markets/jsp/lmp.jsp>

2. LMP is consistent with the actual grid conditions and the dispatch used to keep the lights on.

In the course of calculating spot market prices in real time, the RTO uses the actual generation dispatch and calculates the marginal costs associated with that dispatch. The consistency between the actual dispatch used to keep the lights on and the LMP prices that the RTO pays/charges to the participants in the dispatch helps ensure that those participants have the correct incentives to follow the RTO's dispatch instructions.

Since the LMPs are calculated from the actual dispatch – that is, from the “security-constrained, economic dispatch” that accounts for all transmission network effects -- the

²¹ The information needed for this calculation is readily available from the RTO's dispatch. It consists of the actual dispatch of each generator, the offers/bids of the generators and/or demand-side providers dispatched or available for dispatch, and the actual transmission constraints (congestion, losses, requirements for voltage support, etc) that the dispatch faced during each interval.

LMP prices always reflect real grid conditions, including congestion, losses, the need for voltage support and grid stability and so on. If there are line or plant outages, or unusual events that affect the dispatch, these are reflected in the LMPs. Any grid condition that affects the RTO's dispatch will be captured by the LMP prices.

3. *LMP reflects the lowest as-bid cost of meeting load at each location.*

When calculating the LMP at each location, the LMP processor examines all of the generator offers and demand-side resource bids and determines the lowest cost combination of offers/bids that could serve the increment of load at each location. As a result, the LMPs are not only "marginal cost" prices; LMPs are *the lowest prices at which additional load could be served at each location. LMP is inherently the least-cost approach to spot pricing*; no other pricing method can make that claim.

4. *LMP prices provide the correct price signals to those participating in the dispatch.*

Since the LMPs are consistent with the dispatch the RTO uses to keep the lights on, the prices paid and charged to participants at each location are consistent with the economic incentives each dispatch participant needs to follow dispatch instructions. Or put another way, any price other than the LMP would be inconsistent with the offer prices, and that disconnect would create incentives inconsistent with following the RTO's dispatch instructions. If generators stop following dispatch instructions, the lights would go out.

5. *When an RTO's energy and operating reserve markets are coordinated properly, LMP spot market prices provide the correct incentives to generators about whether to produce energy or stand by for operating reserves.*

LMP captures the tradeoffs and opportunity costs between producing energy versus providing operating reserves. For this reason, the RTO can choose the lowest-cost combination of plants to provide energy and operating reserves *combined*, and the resulting LMP prices will be such that the participating generators or demand-side response providers will be fine with the designations the RTO makes. Thus, if the RTO dispatcher directs Generator A to produce energy instead of standing by in reserve, the

LMP prices for energy and operating reserves will be such that the Generator A will have no regrets from following the RTO's instructions to produce energy. If Generator B is told to stand by with operating reserves, the prices for reserves will be such that it too will have no regrets that it provided reserves instead of producing energy. The LMP price incentives are aligned to support reliability by inducing generators to follow dispatch instructions, producing energy or standing by in reserve.²² Thus, three crucial conditions will hold under an LMP system:

- (1) The RTO will be more likely to acquire the energy and operating reserves it needs to meet all reliability requirements – that is, the probability of keeping the lights on is higher;
- (2) The RTO will get the lowest-cost combination of plants to meet all reliability requirements services – both enough energy and enough operating reserves -- so that the impact on ratepayers will be the lowest consistent with keeping the lights on; and
- (3) No plant or demand-side resource provider will regret the RTO's choice, because in following the RTO instructions, the provider will be better off than if it made a different decision than the RTO's choice.

No other pricing system can meet all of these conditions.

6. The LMP price provides the correct incentives for demand-side responses.

LMP prices reflect the marginal costs of serving load at each location and at each time during the day. So these prices will tend to send excellent signals to consumers (or their utilities or other demand-side providers) about *when* it makes sense to reduce (or increase) consumption and *where* it makes sense to do so.

²² This coordinated approach to selecting plants for energy or operating reserves is sometimes called "simultaneous optimization." *All RTOs agree on the merits of this approach*; different RTOs are at different stages of implementing the optimization rules. For example, NY has a more or less complete set of operating reserve (OR) markets that are simultaneously optimized with the NY ISO's energy spot market; PJM has some OR markets and some optimization, and it is planning to implement remaining OR markets to allow further optimization; MISO is just beginning to develop its OR markets but plans to have a complete set and to apply simultaneous optimization methods.

Because of congestion (and losses), LMPs will be different at different locations. Areas of the grid with plenty of power and/or plenty of transmission to bring cheaper power to those areas will tend to have relative low prices, whereas areas with limited transmission and higher-cost plants will have relatively higher prices; they will have to rely more on local generation that may be more costly.²³ When that occurs, the LMPs will tend to rise in the constrained regions to reflect the scarcity and/or higher costs of local generation and the limited ability to bring in cheaper power from other regions.

Areas with higher LMPs are prime candidates for demand-side efforts by states, utilities, and customers themselves. Times of the day in which spot prices are higher – such as peak demand hours – are also the best times for demand-side reductions to occur. But to recognize these temporal and locational incentives, parties need to see the LMP price differences by time and location as they occur. In other words, those responsible for demand and consumption should face the marginal costs of serving them at each moment in time and at each location. Optimally, this means that retail rates should closely track the hourly wholesale spot prices, for at least those large customers capable of responding to prices

Because the LMPs are immediately posted on the RTO website every five minutes, all parties can know when and where it makes sense to implement demand-side reductions, as well as when/where it is less cost-effective. The LMP data thus becomes a critical set of information for consumers, retailers, utilities, and states in targeting when and where DSM efforts would be most advantageous. Carefully targeted demand-side

²³ Older plants that have higher operating costs tend to be located near urban centers where consumer demand originated and may still be the highest. If there is insufficient transmission to bring lower-cost power into the urban area, it must meet part of its demand from these older, more expensive plants. That will drive up the marginal cost of serving demand in those areas, resulting in higher prices. Note however, these are *wholesale* prices. State regulators have the exclusive jurisdiction to determine how wholesale prices are recovered through *retail rates*, and regulators often adopt retail rate structures that average wholesale costs across a region to create more uniform retail rates across the region. State regulators also tend to fix retail rates at a stable level, but one that recovers the average of wholesale prices. Rate averaging tends to make price signals less effective in encouraging efficient demand-side response, and so states often develop “time of use” rates to better reflect temporal changes in wholesale prices. Larger customers may also be subject to retail rates that vary by location, to reflect locational differences in wholesale (LMP) prices. The incentives to engage in efficient demand-side responses depend greatly on how well retail rates track these temporal and locational changes in wholesale spot prices.

management can increase savings, lower spot prices, lower reserve requirements (and associated capacity and construction costs), improve reliability, mitigate market power, and lower total costs and rates. That is why many states have concluded that improving retail rate structures to encourage more efficient demand-side response is the most important policy tool they have in improving the benefits of the electricity system. And this is true whether that system is served by RTO power pools and related spot markets or exclusively by traditionally regulated utilities. Unfortunately, despite its obvious benefits, this remains a mostly underutilized tool, because regulators are reluctant to allow even larger consumers to face the natural and inescapable volatility of wholesale spot prices.

RTO Spot Pricing Rules Routinely Mitigate the Potential for Market Power

RTOs have no control over who owns generation and cannot force generation owners to divest generation even if they believe generation ownership is unduly concentrated. So the generation in a particular geographical area of an RTO's power pool region may be highly concentrated or highly disaggregated, or anywhere in between. Localized high concentrations of ownership may be the result of a large utility that has not divested its generation or the result of a poorly conceived divestiture, in which much of the generation in a sub-region is now owned or controlled by a single owner or small number of players. However the degree of concentration also depends on whether there are transmission limits that prevent power from surrounding areas to serve load in a given area. Limited transmission into an area can turn local concentration of ownership into a potential market power problem, resulting in excessive wholesale prices.

Although an RTO has no power to force divestiture or change ownership to create more competition, it can encourage transmission expansion to relieve the congestion that creates market power, and it can, through effective spot market pricing, encourage new generation to be built in areas where competition is sparse. These are long-run solutions, but that does not mean RTOs have no ability to prevent the exercise of market power in the short run.

Acting through its Market Monitor, an RTO has substantial authority to mitigate the potential for generation companies to exercise market power in its day-ahead and real-time spot markets. The most important mitigation tools are (1) reducing the prices generators may submit in their price offers for dispatch and (2) requiring generators to submit offers in situations that might otherwise allow generators to withhold their generation from the markets to create artificial shortages that would otherwise raise prices.

By mitigating offer prices, RTOs can limit the ability of any generator to unduly affect spot market prices when the generator is on the margin – that is, when its offer would set or affect the spot market price. An RTO’s market monitor does this by evaluating the level of concentration in a region, taking account of the ability of transmission to bring power from other regions to enhance competition, and using that analysis as a basis for deciding whether to apply offer mitigation and/or must offer rules. The RTO then monitors the price offers that generators submit every day, every hour, and compares them with what the market monitor knows about generator marginal operating and opportunity costs.

When it finds a generator submitting price offers to the RTO market monitor spot market that violate the thresholds for triggering mitigation, the RTO market monitor has the authority to mitigate the price offers – that is, it reduces the offer price and/or assign an offer price in line with the thresholds. The RTO dispatchers then use the mitigated offer price to dispatch the system and to set LMP spot prices. *In other words, offers that may be influenced by the exercise of market power do not set LMP spot prices; spot prices are set from offer prices mitigated in advance by the RTO market monitor to levels deemed reasonable by the market monitor.* Moreover, in all RTOs, the individuals who perform the market monitor functions, including mitigating price offers and must-offer rules, work under strict conflict of interest rules so that there is no connection between them and any market participant. Mitigation thresholds and other market power mitigation rules are filed at and approved by the FERC.

Spot Market Prices Can Provide a Foundation for Recovering Fixed Costs

A common concern about RTO spot markets is that the uniform price auction results in paying market clearing prices that are higher than the marginal operating costs of base-load generators. For example, a nuclear or coal-fired plant may have marginal operating costs that are only \$20 or \$30 per Megawatt hour. At the same time, when these low-cost plants are paid under the uniform price auction, they receive the clearing price in each hour. In any hours other than those with the lowest demand, the clearing price in each hour may be well above the operating costs of these power plants. Is this unjust enrichment?

The answer is no. Every plant must recover not only its operating costs but also its capital (construction) and other fixed costs. The capital/fixed costs of nuclear and coal plants tend to be very high compared to other types of capacity. In a uniform price auction, these costs are not reflected in offers to the RTO; instead the offers tend to reflect only the unit's marginal operating costs. Therefore, every generator must also receive prices above their operating costs during some hours to be able to recover their capital and other fixed costs. In a uniform price auction approach, capital and other fixed costs are recovered through the difference between their operating cost offers and the market-clearing prices they receive from the hourly auctions.

As shown in Appendix Figure 7, market-clearing spot prices may, during off-peak hours, be set by the low operating cost offers of these plants. During those hours, they will recover little or nothing towards their capital/fixed costs. During later hours, as demand rises and market clearing prices rise, these plants begin to receive increasing contributions to their capital/fixed costs. This is shown in Figure 7 by the prices during shoulder and peak hours. And in a system with shortage-cost pricing for operating reserves, there may be a very small number of hours each year in which spot market clearing prices can reach very high levels. During these very high priced hours, all generators receive a significant contribution to their capital/fixed costs.

For a system that has just the right amount of power plants and just the right mix of plants to meet the reliability requirements, the plants with the lowest operating costs will need the total set of revenues depicted by Figure 7. They will receive a small or no contribution during non-peak hours, more during shoulder and peak periods, and more in the rare shortage hours. The combined revenues from all of these hours corresponds to the total revenue requirements for a system in equilibrium – i.e., *one with just the right amount of generating capacity and the lowest cost combination of plants of different types for meeting the reliability requirement*. In that condition, every plant receives the right amount of total revenues, and no plant receives an unjust enrichment.

Equally important, if regulators attempted to prevent these plants from being paid at market-clearing prices (on the assumption they would be *consistently* willing to sell their output at much lower prices), these plants would not recover their fixed costs. Over the long run, fewer of these base load plants would be built, because investors would not invest in projects that could not recover their capital/fixed costs.²⁴

²⁴ This abbreviated discussion describes the long-run equilibrium condition. This paper does not address the merits of how utility generation owners did or did not receive cost recovery under state authorized stranded cost recovery or other transition methods. Nor does it describe efforts state regulators may take to account for market revenues above the costs of rate-based generation. 10/07 -2 jc rev.

Appendix Examples and Illustrations

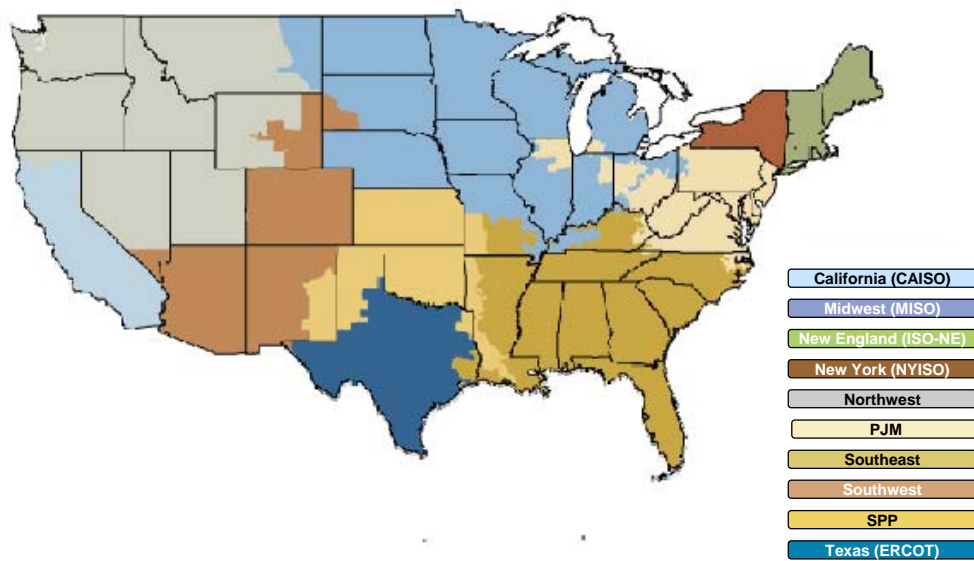
Figure 1. RTOs Operate More Than Half the US Grid

Electric Market Overview

August 2007

Federal Energy Regulatory Commission • Market Oversight @ FERC.gov

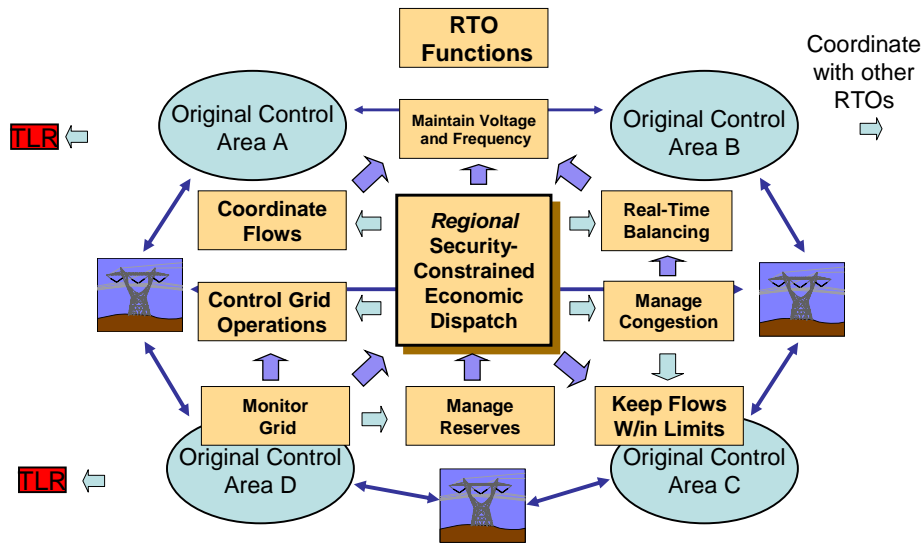
Electric Market National Overview



This paper uses a generic term, Regional Transmission Organization (RTO) to include each of the entities that provides pooled dispatch for a large region via an “independent system operator.” These include the California ISO, Midwest ISO, New England ISO, New York ISO, PJM, Southwest Power Pool (SPP) and ERCOT (most of Texas). All except ERCOT are subject to federal jurisdiction (Federal Energy Regulatory Commission), but ERCOTs dispatch and spot pricing mechanisms are being revised to closely match those used in FERC-regulated RTOs.

Other regions may be served through pool-based dispatch by federal power marketing authorities, include the hydro electric system of the Tennessee Valley Authority, Western Area Power Administration, and Bonneville Power Administration. The remaining regions of the country, particularly large portions of the South and West, are served through dispatches controlled by private and public utilities.

Figure 2.
An RTO Uses a Regional Dispatch To
Replace Local Control Area Dispatches



The core function of any system operator in maintaining reliable operations is the security-constrained economic dispatch. In traditional systems, the utility operates the dispatch for its control area. The dispatch balances supply and demand (plus losses) at every moment, and adjusts the dispatch so that flows across any transmission line are kept within safe operating limits.

In a highly interconnected grid, it is more efficient (less costly) and more reliable to consolidate the dispatch function across several/many interconnected utilities. An RTO does this, consolidating the dispatch function across a wide region to improve reliability and lower costs.

This regional consolidated dispatch feature is the trademark for all “power pools.” An RTO is an “open” power pool, meaning that any generators can sell energy into the dispatch and any loads can purchase energy from the dispatch.

Figure 3.

A System Operator's Dispatch Matches Supply and Demand Every Second

- Dispatchers instruct generators how much to generate at each location in each dispatch interval (usually every 5 minutes).

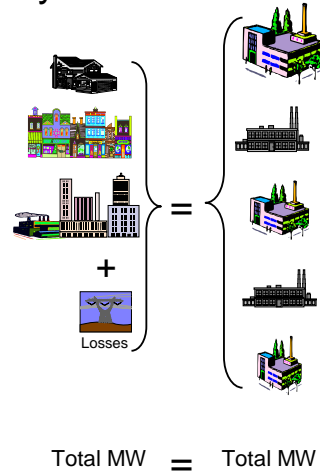
- There's virtually no "storage" in electricity, so electricity must be generated as it is consumed.

- Automated "regulation" fine tunes output in seconds to balance supply/demand at all times.

- Energy dispatch keeps frequency at 60Hz

- Reactive power dispatch keeps voltage stable

- These and other actions keep the lights on

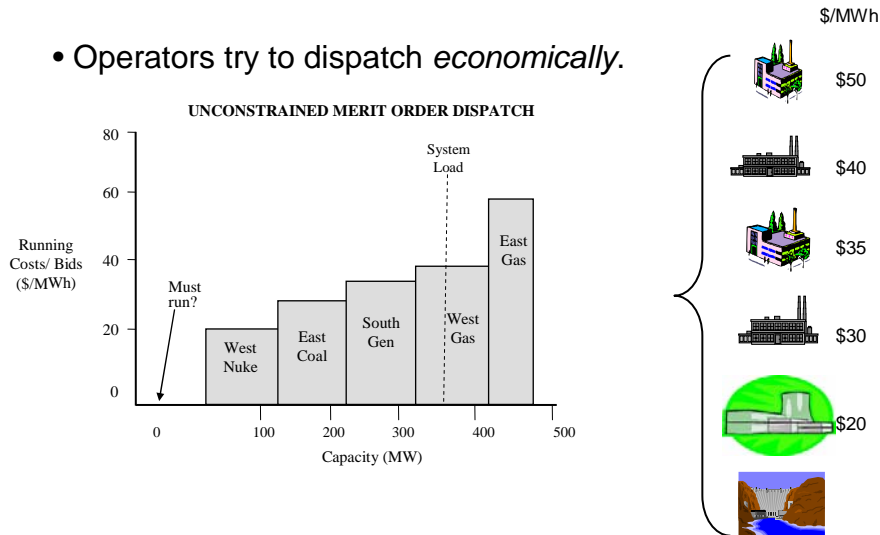


Every system operator, including an RTO, uses the dispatch to keep the system in balance. Because there is no practical way to store large amounts of energy, it must be produced at precisely the moment it is consumed. Production must also cover line losses.

Figure 4.

The System Operator's Dispatch Also Tries to Meet Demand At Lowest Cost

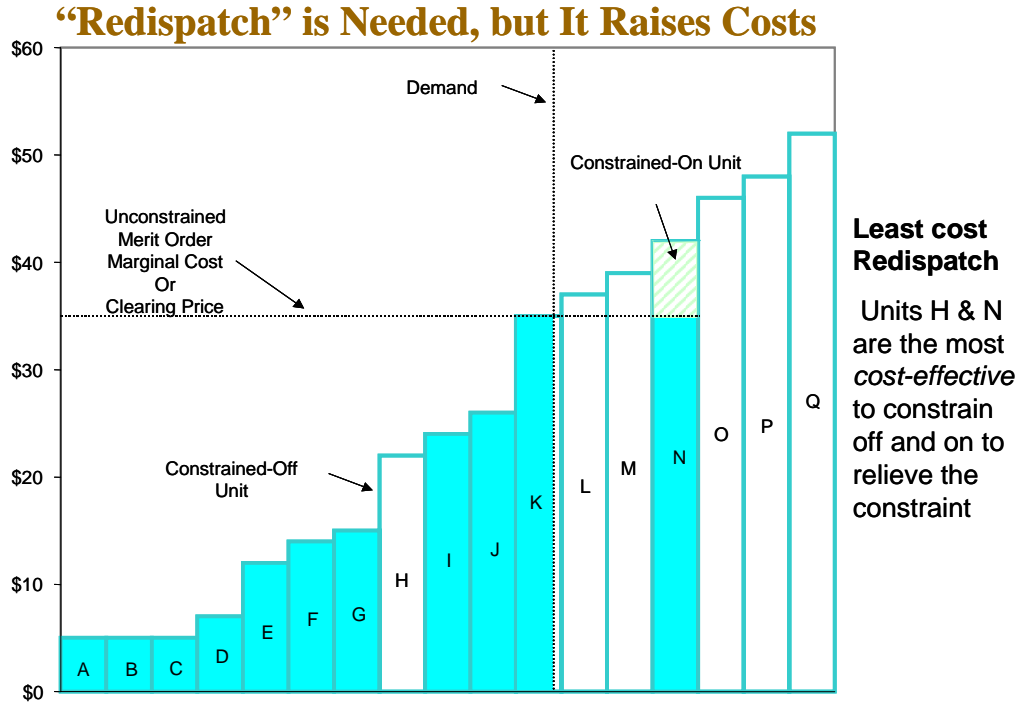
- Operators try to dispatch *economically*.



Every generator has an operating cost, and system operators select the plants with the lowest operating costs (ignores capital or fixed costs) to dispatch to balance the system.

Figure 5.

**Security-Constrained Economic Dispatch:
Congestion Requires Operators to Dispatch Out of Merit Order
to Avoid Overloading Transmission.**



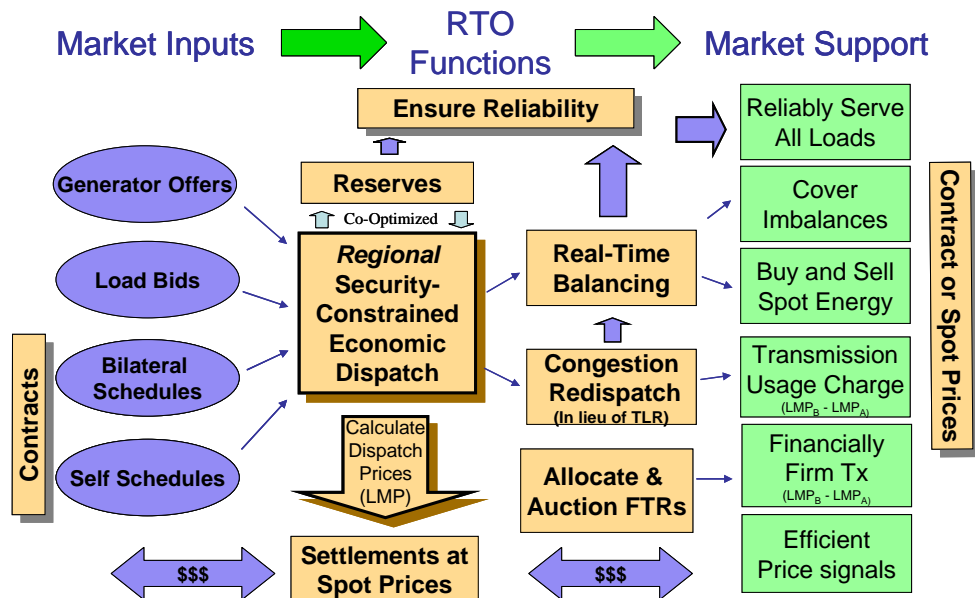
If dispatching in economic merit order would result in flows above security limits for any transmission line or facility, the system operator must change the dispatch to change the pattern of flows. It must constrain off (reduce the output of) one or more plants (H) and constrain on (increase the output) of one or more plants (N). The change in dispatched plants changes the flows across the constrained transmission element. Ideally, the system operator will select the most cost-effective combination of plants to constrain on/off, so as to minimize the total costs of the dispatch while still honoring all transmission safe operating limits.

In the example, the system operator could have constrained off a different plant than H, but because of any other plant's location relative to the constrained transmission, that plant's reduced flows would not have brought flows across the constrained transmission down to its safe operating limit. Similarly, the system operator could have constrained on a different plant than N, but because of location, that other plant's output would not have balanced the load without continuing to violate the safe operating limits of the constrained transmission.

Location and cost both matter in deciding which plants to constrain on and off; the system operator selects the lowest cost combination that still relieves the constraint violation. This is security-constrained, economic dispatch, used by all system operators.

Figure 6. An RTO Operates A Spot Market

RTOs with These Core Features Support Reliability, Renewables, DR and Contracts



The RTO accepts offers from generators and bids from loads and uses these offers and bids to arrange a security-constrained economic dispatch. The dispatch balances supply and demand (plus transmission losses) and is adjusted (“redispatch” – see Figure 5) to ensure that flows resulting from the dispatch do not exceed the safe operating limits on any transmission line or facility.

After arranging the dispatch, the RTO determines the marginal cost of the dispatch at each location on the grid. The resulting locational marginal prices (LMP) become the spot market prices that are used to pay generators who sell energy via the dispatch and to charge loads that purchase energy via the dispatch. In this way, the financial settlements for buying and selling through the dispatch automatically define an electricity spot market. Imbalances or deviations from schedules can then be paid/charged at the LMP.

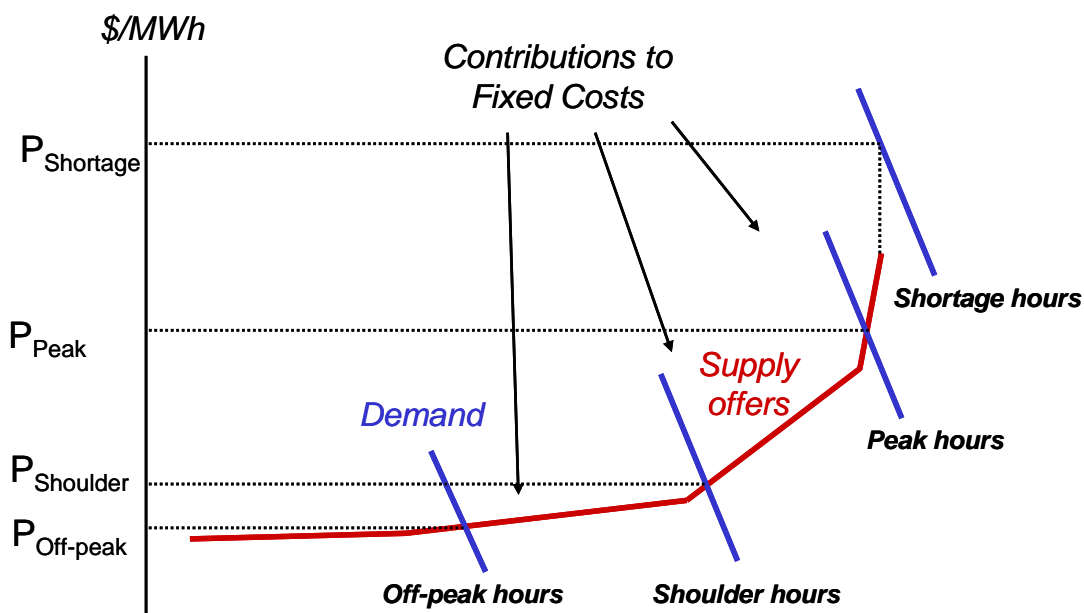
Generators are paid the spot market price (LMP) at their location for their energy.

Loads pay the spot market price (LMP) at their location for the energy they buy.

Parties who self schedule their own generation or schedule contracts indicate injections at one location and withdrawals at another location. Schedules pay for any redispatch required by their grid usage. The redispatch price is the difference between the LMP at the receiving location (B) and the LMP at the sending location (A) times the amount of MW scheduled or transmitted.

Figure 7.

Generators Depend on the Highest-Price Hours To Recover Most of Their Fixed Costs



Low-price hours barely cover operating costs

As demand for energy increases throughout the day, the increasing demand curve crosses further up the supply curve of generator offers, raising the clearing price.

During off-peak hours, only plants with the cheapest operating costs (offers) are needed to meet demand, and so the clearing prices remain quite low. No plant, not even the least expensive to build, can recover its capital/fixed costs at that price.

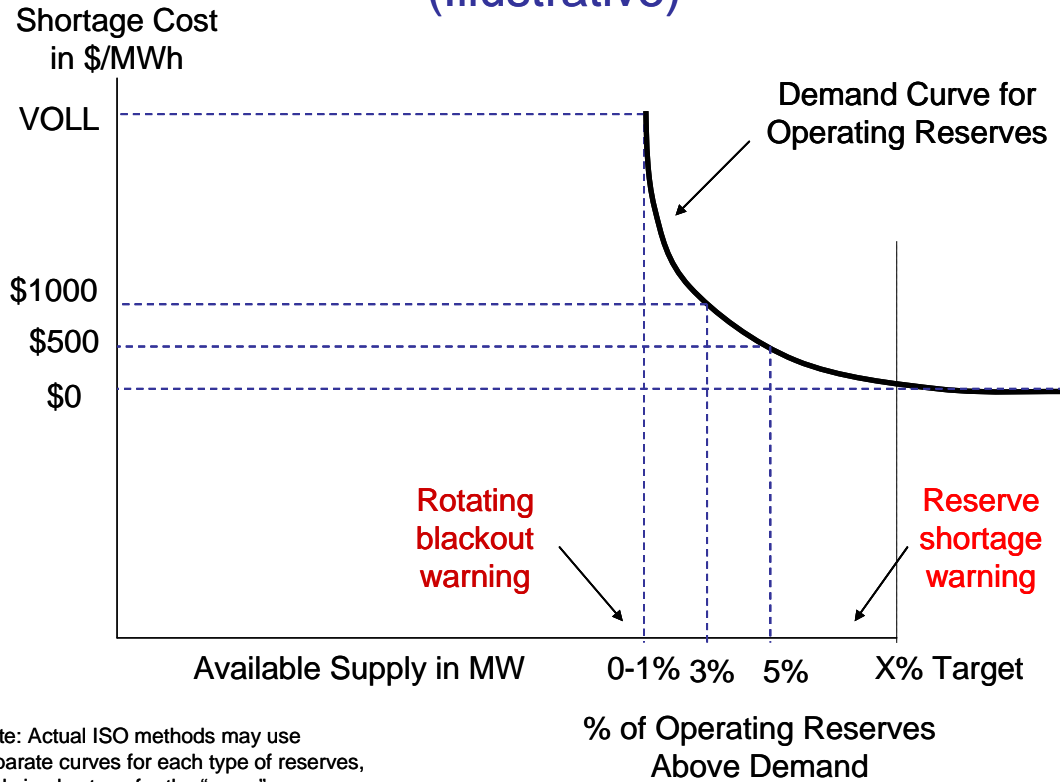
As the price rises later in the day, higher prices allow plants to begin recovering their capital/fixed costs. However, note that even when prices are high during peak hours, peaking plants would recover little towards their capital/fixed costs, and other plants would recover something, but not enough revenues.

Only prices set by “shortage-cost pricing” during the very few hours when the RTO is short of operating reserves can complete the recovery of capital/fixed costs for all plants.

Currently, only NY-ISO has shortage-cost pricing with operating reserves, and even they require a capacity payment to complete cost recovery. Other RTOs cap prices below shortage-cost levels; they still need capacity markets and capacity payments to provide the revenues needed to recover capital/fixed costs. Without those payments, investors would not build all the plants needed to meet the reliability target (e.g., X% reserve margins) because they would not receive a sufficient return of or on their investments.

Figure 8.

Shortage Cost Pricing Approach (Illustrative)



In this illustration, the RTO creates a “demand curve” reflecting the effect on energy prices as the RTO becomes increasingly short of its target level of operating reserves.

When the RTO has adequate operating reserves, then there is no “cost” associated with reduced reliability; hence the impact on the energy price is \$0. However, if the RTO has less than the target level of reserves, say only 5 percent when it’s target was higher, then the curve defines an addition to the energy price – in this case \$500/MWh.

As operating reserves approach 0-1 percent, at which curtailments might be imposed (rotating blackouts), the price of energy rises towards an average “value of lost load,” which represents the average price at which non-price-responsive consumers would be indifferent to being served or curtailed.

The effect of using a shortage-cost pricing approach is that it (1) reflects the value/costs of shortages well before curtailments become necessary and therefore (2) strongly encourages more generation to come on line and/or loads to engage in demand response (DR), well before actual curtailments occur. The combined effect is to enhance efforts to keep the lights on, reduce shortages and the need for curtailments, and limit prices. Enhanced DR also has the benefit of mitigating attempts to exercise market power.