Chapter 5
The Design of US Wholesale Energy and Ancillary Service Auction Markets: Theory and Practice

UDI HELMAN,[1] BENJAMIN F. HOBBS,[2] AND RICHARD P. O'NEILL[1]

[1] *Federal Energy Regulatory Commission, Washington, DC, USA;*
[2] *Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, MD, USA*

Summary

In the United States, after about a decade of experience with market design, wholesale spot markets operated by Independent System Operators (ISOs) around the country have largely converged on core design elements. This chapter provides a detailed description of how these markets operate. In particular, most of these markets have day-ahead and real-time auction markets for energy and certain ancillary services, typically regulation and operating reserves. The energy auctions can accommodate both physical and virtual supply offers and demand bids. In the regulation and reserve auctions, only physical offers are currently allowed, including those from dispatchable demand. With the submitted day-ahead offers, bids, and non-price schedules, the ISO conducts a security-constrained unit commitment auction, which selects the generation units that will run for every hour of the day subject to all relevant units and transmission network constraints. For the energy markets, the auction outcome is two sets of prices that together clear the market: locational marginal prices (LMPs) for energy, which include congestion and loss components, and separate payments to ensure revenue sufficiency for any offer or bid costs, such as generation start-up costs, not recovered through LMPs. The real-time energy markets have also progressively incorporated most elements of this design, although auction procedures are somewhat different from the day-ahead market.

Integrated into the day-ahead and real-time energy markets are markets for regulation and operating reserves, co-optimized with energy. Market prices for these ancillary services typically incorporate an opportunity cost payment with respect to any foregone energy sales as well as availability payments, if needed. As with energy, revenue sufficiency is guaranteed through additional payments. To provide insight into each stage of the market and into the principles of locational marginal pricing, the chapter provides

a simple numerical example of an energy auction on an electricity network. Finally, the chapter briefly explores other key design issues, such as refunds of surplus marginal congestion and loss payments, market power monitoring and mitigation, addressing continuing market seams, software development, and extensions of the market design.

5.1. Introduction

In the UNITED STATES, wholesale markets for electric power have evolved along two basic organizational approaches, both consistent with the open access transmission regime established by the US Federal Energy Regulatory Commission (FERC) in 1996 (FERC, 1996a). In the first type of market, electric utilities and non-utility generators contract bilaterally among themselves for energy on a forward basis. The utilities that own the transmission facilities determine the available quantity and price of transmission access and physical scheduling rights, subject to open access rules. While private power exchanges may form to facilitate forward energy contracting, there is no co-ordinated spot energy market that encompasses the territory of multiple utilities and heretofore no price-based congestion management. In the second type of market, an independent third-party entity, which for purposes of this chapter will be called an Independent System Operator (ISO), operates organized regional bid-based auction markets for spot energy, various types of ancillary services, and possibly capacity, and allocates all transmission capacity and transmission property rights in an efficient and non-discriminatory fashion. Spot transmission usage is subject to charges for congestion and losses. These ISO markets are the subject of this chapter. The chapter will focus on the design of the daily energy, regulation, and operating reserves markets. The chapter will not review ISO market performance, except on occasion to explain a design decision.[1]

5.1.1. Overview of Market Design

The ISO market designs that have arisen in different regions of the United States have similarities and differences, resulting from the fairly high degree of regional decision-making in the regulatory reform of the US electricity sector.[2] In general, the ISO operates a day-ahead market and a real-time, or dispatch, market. These are organized as sealed-bid, multiple-unit auction markets ~~with uniform market clearing prices~~. The day-ahead market is a forward market in which accepted offers or bids can choose not to perform in real-time (i.e., go to physical delivery) as long as they buy back or sell back their positions. Put another way, the real-time market determines the prices of “deviations” from the day-ahead schedule. Due both to financial incentives and certain administrative rules discussed below, most accepted day-ahead offers and bids that reflect physical supply and demand do go to physical delivery and hence the two markets collectively can be thought of as the “spot” market.

In both markets, suppliers submit offers prior to a trading deadline (usually the prior morning for the day-ahead market and about 1 hour before the real-time market). These

[1] Market performance is discussed extensively in the US ISO annual state of the markets reports, e.g., PJM 2000–2007, Potomac Economics 2003–2006 (for New York ISO). In addition, there is a large research literature on this subject.

[2] The evolution of US federal rulemaking, including the failed effort to standardize the market designs, is discussed in O’Neill et al. (2006).

offers must usually specify the minimum and maximum MW that can be produced by the generator, the price of energy ($/MWh) over the range of its available output, a start-up cost ($), a no-load cost ($), and a number of physical characteristics, such as how rapidly the generator can increase or decrease output (called the "ramp rate" and measured in MW/minute). For operating reserves and regulation, suppliers typically provide additional price offers and physical parameters to define capability.

Both the day-ahead and real-time auctions conduct "unit commitment," in that they specify exactly which generation units should be turned on in each hour, their level of output, and the length of time they should run over the day, based on start-up and energy offer prices and the other financial and physical parameters. In the day-ahead market, the unit commitment decision is integrated into the auction. In real-time, unit commitment is conducted through a parallel pre-auction program that "looks-ahead" based on a forecast to determine which units to commit or decommit, while the auction function only adjusts the output of already committed units on a 5–15 minute basis. In addition, because energy, regulation, and reserves can be either or both complementary and substitute uses of a generator, establishing the correct auction constraints to reflect these possible relationships, both day-ahead and real-time, has proven to be important for purposes of economic efficiency.

Please move these 2 sentences, with edits (deletions and inserts), HERE

A key function of the energy auctions with LMP ~~in the US ISOs~~ is to put a market price on marginal transmission usage, including congestion and losses. In the day-ahead market, congestion is managed instantaneously as part of the auction optimization, which respects most relevant transmission constraints. In the real-time market, the system operators manage congestion on a minute-to-minute basis in part through auction prices and in part through non-market operating decisions. For both markets, the auction result is a market ~~equilibrium~~ with uniform ~~market clearing~~ prices at each specified commercial location (nodes or buses) on the transmission network, called "locational marginal prices" (LMPs). Buyers pay the LMPs at their locations (which to date have typically been zonal prices comprised of a load-weighted average of the LMPs in the zone) and sellers are paid the LMPs at their locations. Regulation and operating reserve prices are calculated slightly differently, typically incorporating an opportunity cost payment and an availability price offer into the calculation of market clearing prices, which are determined for pre-defined zones. Additional "pay-as-bid" payments are made to accepted offers and bids in any of these markets if their auction revenues do not fulfill their offer or bid terms (e.g., if a generator is started up but then does not run for sufficient hours to cover the sum of its offer prices for start-up, no-load, energy, regulation, or reserves).

These vast regional wholesale spot markets, several consisting of tens of thousands of simultaneously determined prices at locations on the grid, are one of the signal technological achievements to date of the regulatory reform of the US electricity industry. The ISO market designs are also serving as models for other countries' attempts at market development. At one level, they operate very well, allowing for the regional market operator to capture efficiencies made possible through large-scale optimization. Buyers can schedule their own resources (owned or contracted) or buy spot. Sellers can optimize whether to fulfill their forward contracts with their own resources or through the spot market and can offer any residual capacity into the spot market.

The record of the first decade of energy and ancillary service market design in the US ISOs suggests that while some conceptual design issues have largely been resolved, and there have been many technological innovations and advances in auction design, there remain many design and implementation challenges. First, these markets remain "incomplete" in the sense that they do not price adequately all generator services and physical requirements associated with power transactions (transmission markets present other types of incompleteness). Moreover, while sellers have sufficient flexibility to offer their true costs, buyers are not sufficiently price responsive, due primarily to regulatory and technological barriers. Second, the ISO markets, while fairly unconcentrated in the aggregate, are not perfectly competitive at all locations and at all times due to transmission constraints. The primary method to control intermittent market power in the ISO markets has been through supply offer price caps. But offer caps suppress market prices and sometimes hinder investment. As a result of these two problems, additional pricing rules, such as the "scarcity pricing" discussed in this chapter, and additional markets established through regulatory requirement, such as capacity (or resource adequacy) markets, have been perceived as necessary to support both power system reliability and economic efficiency. Some of these design features can be removed (or become irrelevant) as technology evolves and the markets become more complete.

With respect to several design elements, no definitive best practice has yet emerged. To illustrate this, in several sections of the chapter, the rules and procedures in two US ISO markets, PJM and New York, are discussed. These markets were picked because they offer useful comparisons on several design choices in the auction market rules, the integration of market and system operations, and market power mitigation. Nevertheless, the differences, while worthy of consideration, should not obscure the convergence of the US ISO markets over time on important common design elements.

5.1.2. Chapter organization

In an earlier survey (O'Neill et al., 2006), electricity market design principles were discussed at a relatively high level and US ISO market features were reviewed. The complementary purpose of this chapter is to provide an in-depth description of the sequence of the daily ISO auction markets for energy and certain ancillary services. There is sufficient detail to allow the reader to track most of the market rules and computational procedures that affect the ultimate price of wholesale power transacted through such markets.

The chapter is organized as follows. Section 5.2 provides an overview of electricity market design choices and options for market power mitigation. Sections 5.3–5.5 discuss the auction sequence of the day-ahead market, the reliability unit commitment, and the real-time market. A numerical example is introduced in Section 5.3 to motivate the auction description, and continues in most subsequent sections. A mathematical statement of the auction model used in the example is provided in ~~the~~ appendix 5A. Section 5.6 examines revenue sufficiency guarantees that support efficient market behavior in each step of the auction sequence. Section 5.7 explains how surplus spot energy payments collected by the ISO due to pricing of congestion and losses are refunded to market participants. Section 5.8 discusses market power monitoring and mitigation in these auctions. Section 5.9 collects several additional topics in ISO market design and implementation, including the ISO's longer-term markets and operational functions. Section 5.10 discusses possible next stages in the design of these markets. Section 5.11 presents conclusions.

5.2. The Development of Wholesale Energy Auction Market Designs

The designs of the wholesale auctions for energy and ancillary services described in this chapter emerged over a decade of research and experience in the United States and elsewhere and continue to be modified and refined [see, e.g., the surveys in O'Neill et al.(2006); FERC (2002); Stoft (2002); Wilson (2002)]. This section briefly reviews some of the alternative designs that were considered in that period and explains why certain choices appeared desirable for theoretical, practical, and regulatory reasons. These include

- definition of the market products,
- functions of the ISO,
- choices in electricity auction design, including the question of uniform versus discriminatory pricing,
- sequencing of markets and reliability functions,
- market power monitoring and mitigation, and
- scarcity pricing.

5.2.1. Definition of market products

A first step in designing markets is to define the products, including the time and location at which the transaction takes place. While in most commodity markets, product definition is a matter for private firms to determine, in the electricity spot markets operated by ISOs, the nature of the technology in real-time (e.g., balancing requirements and lack of storage) and the presence of reliability requirements has led to a strong regulatory role in product definition (e.g., in FERC, 1996a), although ISOs have had latitude to adjust some definitions and parameters to fit their systems.

Energy is the primary wholesale product traded in the ISO electricity markets (as measured in terms of quantities and monetary value), and it is defined straightforwardly as mega-watt-hours (MWh) injected or withdrawn at a location or locations (e.g., hub or zone) on the transmission network in one of the hourly markets (day-ahead or real-time). For purposes of spot energy market design, sales of energy are typically limited to generator output between a unit's minimum and maximum operating levels. Price offers for energy are typically required to be ~~continuous and~~ linear (see Table 5.7). Generator "start-up," i.e., the short-term fixed costs associated with accepting a generator offer that requires a unit to start-up, is in a sense treated as a separate, discrete product with its own pricing rules (under the revenue sufficiency payment), as is operating the generator at ~~the~~ "no-load" level. A parallel construction is found on the demand side, in which the demand may bid to consume MWh along with short-term fixed costs associated with implementing a demand curtailment offer. The pricing and settlement rules are discussed further in this section as well as extensively in Sections 5.3–5.6.

The ancillary services are reliability services offered by eligible suppliers or responsive demand. As summarized in Table 5.1, these include regulation and different types of T1
operating reserves, measured both as energy and as capacity (MW) made available at locations on the transmission network in one of the hourly markets (day-ahead or real-time). The spatial aspect of regulation and reserve product definitions is typically slightly different from that of energy. The ISOs define locations for these products on a zonal basis. ~~This is because they are not provided on a bilateral basis to particular buyers, but instead are made available to the aggregate of buyers in the part of the system to which they are deliverable.~~ These products share the characteristic that they are difficult to disaggregate among each buyer on the system, so they are currently procured by the ISO on behalf of buyers, who pay a load-weighted average price. Because regulation and spinning reserves are complementary or substitute uses of generator providing energy (or a load consuming energy), their definition typically incorporates an energy component. The implications for auction design are discussed below.

Table 5.1. Definitions and characteristics of market-based ancillary services in the US ISOs currently in operation or under consideration

Type of ancillary service	Description	US ISOs with markets
Regulation (or automatic generation control, AGC)	The ability to increase or decrease energy output on a second-by-second basis for energy balancing	New York ISO, ISO New England, PJM, California ISO
Ten-minute spinning (or synchronous) reserve	Reserves available (MW) within 10 minutes from generators synchronized with the grid or demand response	New York ISO, ISO New England, PJM, California ISO
Ten-minute non-spinning or non-synchronous reserve	Reserves available (MW) within 10 minutes from generators not synchronized with the grid or demand response	New York ISO, ISO New England, California ISO
Thirty-minute or supplemental reserves	Reserves available (MW) within 30 minutes or more from generators either synchronized or not synchronized with the grid or demand response	New York ISO, ISO New England, California ISO
Reactive power	A product of generators and types of transmission elements that essentially supports the voltages that must be controlled for reliability	Currently cost-based procurement; markets under consideration (see, e.g., FERC, 2005).

5.2.2. Functions of the ISO

A major market design question in the mid-1990s concerned the relationship between the transmission system operator and the operator of the spot markets. This issue takes on different forms in different countries, due to the requirements of their regulatory systems and the history of the electric industry. For example, in some countries (such as England and Wales), the transmission system operator is a single firm under performance-based regulation (see, e.g., Chapter 4). In the United States, there was no such opportunity to merge the ownership of national or regional transmission operations at the start of restructuring. Instead, there has been an ongoing experimentation on a regional basis with different organizational approaches to the implementation of transmission open access and management of transmission usage. And as a result, options for designing the energy and ancillary service markets also take different forms in different regions of the United States.

As noted in Section 5.1, there are currently basically two organizational approaches to transmission system operations in the United States: individual utility operations and ISOs. To understand the purpose behind the ISO market, it is worth beginning with a description of markets where the grid is operated by individual utilities. In such markets, all energy transactions (long-term and short-term) are made by individual buyers and sellers (i.e., bilaterally) but there is no central trading point that co-ordinates transmission usage. Hence, before or after an energy contract is in hand, the buyer or seller must arrange a transmission contract for a pre-specified period that matches the energy contract's source point(s) and sink point(s), and contains other characteristics such as the priority given to the transmission reservation (sometimes called the "firmness" of the transmission reservation).[3] Ancillary services, including balancing energy, are bought under cost-based rates (FERC, 1996a). This type of market has supported an increase in wholesale energy transactions, especially with the advent of the Internet (FERC, 1996b). Its main limiting factors as a market design are the following:

- The lack of co-ordinated transmission operations to efficiently utilize transmission capacity across multiple utilities, especially when there is congestion (FERC, 1999, 2002). AU2
- The lack of a centralized exchange or auction to facilitate spot trading of energy or ancillary services (FERC, 2002).

A higher degree of market organization, as offered by an ISO, is thus generally desirable to support market efficiency and expand market scope.

Turning first to organization of transmission operations, market improvements can be achieved in several ways, with different implications for market efficiency. First, individual (transmission-owning) utilities can retain control of their transmission systems, but can establish an independent entity (sometimes called a "Day 1 ISO") to facilitate information exchange and transmission capacity reservation by parties seeking to buy transmission. This type of organization can improve transmission scheduling to facilitate energy trading.[4] Second, individual utilities can cede full ~~operational~~ control over their transmission systems to an ISO, which will simultaneously operate the entire transmission network collectively along with real-time energy and certain ancillary service markets. Whether the ISO also operates energy and ancillary service markets prior to real-time, such as day-ahead markets, has been another design decision.

Whichever type of centralized transmission operations is chosen, an organized market for electricity can be developed along several different formats. If that market is going to be organized at a central location, then it is likely to be either an exchange or an auction. An exchange is a common format for commodity forward and futures trading. This is typically an "open" market in that buyers and sellers are known to each other. The most common pricing rule is the "bid-ask" method characteristic of futures exchanges. In an auction, a third party known as an auctioneer matches buyers' and sellers' following a pricing rule. Either an exchange or an auction could be used to organize spot electricity markets, but for reasons discussed next, an exchange is more compatible with pre-day-ahead forward markets while an auction is best suited to the day-ahead and real-time spot markets operated by ISOs.

5.2.3. Choices in electricity auction design

The electricity markets described in this chapter are all auction markets. To expand on the description given above, Table 5.2 summarizes some common auction designs (see, e.g., Klemperer, 1999; Krishna, 2002). In general, the choice among the designs depends T2

[3] In the US context, firmness designates the physical priority of a point-to-point transmission contract in the event that reliability concerns, such as unmanaged congestion, prompt the system operator to scale back ("curtail") transmission usage. Typically "non-firm" contracts, which are usually shorter-term, are curtailed before "firm" ones. In the United States, this arrangement prevails only outside the ISO markets, which use price-based congestion management rather than physical priorities. See also discussion in Chapter 4.

[4] The Midwest ISO operated as such a transmission scheduler from 1997 to 2005, when it began operating a market with the design described in this chapter.

Table 5.2. Comparison of some alternative auction designs

	Auction type	Description	Revenue adequacy	Incentive compatibility
Single-Unit Auctions	Open ascending price ("English")	A one-sided auction in which the auctioneer raises the price of the unit being sold until there is one remaining bidder. The selling price is the winning bid.	Yes	Yes. The bidder with the highest valuation would remain ~~until the price exceeded its own valuation and the winner would pay the value~~
	Open descending price ("Dutch")	A one-sided auction in which the auctioneer lowers the price of the unit being sold until one bidder is willing to purchase. The selling price is the winning bid.	Yes	~~No.~~ The bidder with the highest valuation for the item would have the incentive to wait past where its value is in order to receive a lower price.
	Sealed-bid first price	A one-sided auction with sealed bids in which the selling price is the highest bid price.	Yes	No. The bidder with the highest valuation would want to shave its bid to receive a lower price and thus could lose out on winning the auction
	Sealed-bid second price	A one-sided auction with sealed bids in which the highest bid wins the auction, but pays the bid of the next highest bid.	Yes	Yes. Each bidder would have the incentive to submit its own valuation.
Multiple-Unit Auctions	Sealed-bid uniform price	A one- or two-sided auction with sealed bids in which the market clearing price for all units is the highest bid price for any unit.	Yes	Not completely. If the bidder can influence the price, it would have the incentive to buy less in order to lower the price for the amount it buys
	Sealed-bid discriminatory (pay-as-bid)	A one-sided auction with sealed bids in which each seller that clears the auction is paid its offer price for any individual unit.	Yes	No, the bidder would have the incentive to guess what the clearing price would be.
	Sealed-bid Vickrey		No	Yes, each bidder would pay the opportunity cost imposed on the other players if the bidder had not been in the market, and as such would have the incentive to bid its true valuation.
	Open auctions (English or Dutch)	These are similar in design to the single-unit English and Dutch auctions.	Similar to single unit auctions.	Similar to single-unit auctions.

Sources: Krishna, 2002; Hobbs et al., 2000.

on a number of factors, including the characteristics of the product being auctioned and the revenue and efficiency properties of each ~~design~~. The typology in Table 5.2 divides auctions into either "single-unit," in which buyers and sellers are trading single items or single bundles of items (such as a single MW or a "strip" of MW over multiple hours), or "multiple-unit," in which multiple units are traded simultaneously.[5] As noted, these types of auctions can then be further divided into open or sealed bid formats. Under the sealed bid format, buyers and/or sellers submit bids and offers that are not known to others. The anonymity of a sealed bid auction is further compounded in the spot electricity auction, in which buyers do not know which seller they are buying from. Within the multiple-unit, sealed-bid auction framework, the two major choices for pricing are discriminatory, or "pay-as-bid," prices or uniform clearing prices.

Because auction designs must fit the commodity being traded, not all the auction designs shown in Table 5.2 are applicable or desirable for the electricity commodities discussed in this chapter. The product definitions and the technological characteristics of wholesale electricity vary in the different stages of the market: the pre-day-ahead forward markets that take place outside the ISO markets, where trading is primarily for multi-day/multi-hour strips of power; the day-ahead markets, where trading is on an hourly basis; and the real-time spot markets where power goes to physical production and delivery and trades takes place within the hour itself. The clearest way to describe how auctions can be designed around these forward and spot markets is to begin with the ISO's real-time market and work backwards in time.

The real-time market has a number of characteristics that greatly narrow the alternatives for auction design, at least with current power system technology. In real-time, the system operator is balancing supply and demand on a time-frame of seconds and minutes, largely by directing the output of generators on the system over such time-frames (and some aspects of the transmission facilities). Because of the nature of power flows, the injections and withdrawals of power on the system must be balanced simultaneously; i.e., individual injection and withdrawal combinations cannot be evaluated independently. Hence, there are no physical "bilateral" transactions involving matching between buyers and sellers in real-time: all purchases are either via the ISO market or self-provided.[6] The real-time dispatch of generators must also account for inter-temporal constraints, such as ramp rates, that create interdependencies across hours. Moreover, demand in real-time can be price-responsive, but typically not on a time-frame of minutes or seconds; hence, such demand will inevitably remain largely price-inelastic until technology provides greater responsiveness and the number of units of electricity (e.g., MW) consumed will fluctuate from minute to minute.

These factors inevitably require that a real-time auction is multiple-unit and sealed-bid and that the market ~~is~~ simultaneously cleared with all relevant transmission and generation constraints taken into account. That is, it is simply not possible to have a single-unit electricity auction in real-time, even for multi-hour strips (although that can be done in the forward markets) and it is hard to conceive of an open auction format that can meet the time-frames necessary. Essentially, real-time electricity auctions as they currently function in the United States are simply least-cost ("economic") security constrained dispatch

[5] Multiple-unit auctions do not have to take place simultaneously, but could instead take place in a sequence of single-unit auctions (Krishna, 2002); however, this latter format is not consistent with the physical characteristics of electricity.

[6] Prior to the formation of ISOs, individual utilities provided other parties that were using their transmission facilities with "energy balancing," typically priced on an average basis.

using market-based supply offers and demand bids to determine prices. These auctions will be described in more detail below. Whether the pricing rule should be uniform or discriminatory will also be discussed in the next section.

In the forward, but short-term, electricity markets, more auction design alternatives become possible, including some simplifications of ~~the~~ commodity pricing, although still constrained by their proximity to the physical requirements of the operating day. For close to real-time auctions, such as hour-ahead or day-ahead auctions, deviations from the auction design for the real-time market are possible, such as relaxing some of the physical constraints on the network, but consideration must be given to implications of such designs for reliability and economic efficiency. Economic efficiency will be adversely affected if the forward auction schedule – i.e., the scheduled output of generators going into the operating day – is substantially different from the real-time dispatch, and if the resulting adjustments for dispatch feasibility incur otherwise avoidable costs or if such costs are assigned on a basis that undermines efficient price signals.

For example, in the early phases of ISO auction design in the United States, some parties preferred day-ahead auctions or exchanges that did not meet the requirement of simultaneous feasibility of the resulting day-ahead schedule using an actual network. Instead, they argued for a zonal approximation of the network to facilitate day-ahead trading. This was the design choice adopted in the first phase of the California market (1998–2001), in which there was a separate day-ahead market operated by the California Power Exchange (PX) and a real-time market operated by the California ISO. To facilitate the PX market, California was divided into two transmission zones that largely, but not exactly, corresponded to the major transmission constraints on the system. The PX operated a single market in each zone using a multi-unit auction with a single zonal clearing price (see, e.g., Sweeney, 2006).

The resulting schedule was then passed to the ISO, which could aim to adjust it to the limits on inter-zonal transmission through adjustments offers and bids available from the PX, but which could not enforce an efficient schedule by altering the scheduled output of generators on the basis of their supply offers. In real-time, the ISO would have to make any final scheduling adjustments on both an inter-zonal and an intra-zonal basis using real-time energy and regulation offers. In concept, the PX was to be joined over time by other competing day-ahead ~~“scheduling co-ordinators,”~~ which could all operate their own auction markets or exchanges under whatever design suited them. The ISO would thus have been collecting many day-ahead schedules and attempting to impose feasibility on them in a similar fashion. Over time, evidence collected that, along with creating cross-subsidies on an intra-zonal basis, such zonal pricing did not provide appropriate locational signals for investment in generation, transmission, and demand response. California ISO will shortly revise the market design to adopt the day-ahead and real-time auction with locational marginal pricing described in this chapter.

Like energy used for balancing, the real-time markets for regulation and operating reserves are most compatible with the sealed-bid, multi-unit auction design with all network constraints represented and uniform prices. Again, day-ahead markets for these services could use other auction designs but, as with energy, potentially resulting in market inefficiency. For these products, the primary design issues have been (i) the sequence in which the energy, regulation, and reserves markets are cleared; (ii) whether they are co-optimized or not; (iii) how complementarities and substitutions between them are captured in the auction pricing algorithm; and (iv) the components of the supply offers. These design issues have been reviewed in FERC (1999a, 2002), Stoft (2002), and O’Neill et al.(2006), among others. ~~For purposes of this chapter,~~ the US ISOs that offer these ancillary services through auctions have largely converged on a design that incorporates co-optimization and represents a hierarchical substitution that supports an efficient use of generators and any demand resources that can provide them. More detail on these auctions follows in Sections 5.3 and 5.5.

5.2.3.1. Uniform versus discriminatory (pay-as-bid) pricing

As noted, within the multiple-unit, sealed-bid auction format, there are two primary pricing alternatives: uniform or discriminatory, also known as pay-as-bid (Klemperer, 1999; Krishna, 2002). At present, there is no theoretical consensus on the revenue and incentive compatibility properties of these rules as applied to electricity auctions (see, e.g., Kahn et al., 2001; Fabra et al., 2004). However, there is agreement on their implications for market transparency and also market power mitigation, as described later. Those implications have made the uniform pricing rule more attractive in the US ISO markets (although the pay-as-bid rule is currently used in the England and Wales spot market). There are also variations within each type of pricing rule that have implications for market efficiency.

Under uniform pricing, all sellers are paid the price offered by either the last unit (MWh) chosen by the auction or the next (incremental) unit that is not chosen (the exception to this rule in some ISOs is in periods of scarcity pricing, as discussed shortly, during which the market price is set through a demand curve). For example, if a $40/MWh offer is accepted and the next offer in the offer stack that is not accepted is for $50/MWh, then the clearing price for all sales is set at either $40/MWh or $50/MWh. In the real-time market for energy, the choice between these alternative uniform pricing rules has to do with how the ISO seeks to control the output of the marginal generator(s) through a combination of price signals and quantity instructions. The choice also has an impact on long-term pricing signals. For example, if due to generation constraints, such as generator ramp rates (and assuming no transmission congestion), the ISO has to turn on a $50/MWh unit while a $40/MWh unit is operating at below its economic maximum output, but that the next available MWh is at the $40 price, then the choice of pricing rules has clear incentive properties:

- If the market price is set at $50/MWh, then the $40/MW unit may have an incentive to increase its output. If this creates reliability problems for the system operator to manage, then it will have to control the output of the $40/MWh unit through dispatch instructions and penalties for deviations (as is done in many ISOs).
- If the market price is set at $40/MWh, then the $40/MWh unit has no incentive to increase output (assuming that its offer is reflective of its marginal cost), but an alternative pricing rule has to be established to pay the $50/MWh unit to turn on. Typically, the unit is paid its bid and the ISO collects the revenues from buyers.

On an electrical network with nodal congestion pricing, this auction pricing rule results in uniform clearing prices at each location and is called locational marginal pricing. A more detailed example of this pricing rule is provided in the numerical example that begins in Section 5.4.

~~An alternative auction pricing rule, that has been discussed in the United States and elsewhere, is to pay each accepted supplier individually at its offer price (e.g., discussion in Kahn et al., 2001). This is called discriminatory or pay-as-bid pricing.~~ In pay-as-bid designs, the supply offer is typically conceived as a bundled offer for energy, start-up, and no-load (sometimes called a "one-part" offer). For example, if a supplier makes an offer of $20/MWh and is cleared through the auction, then the offer is paid $20/MWh, no more or less. Under most proposed formats, buyers in the auction would pay an averaged price based on the total accepted MWh. Hence, while a locational pricing result could be sustained for sellers, it would be difficult to sustain for buyers, except on an aggregate basis, such as in a zone.

Advocates of pay-as-bid pricing have different motivations, depending on whether they are sellers or buyers. Sellers may believe that the combination of one-part offers with pay-as-bid provides the seller greater transparency over the auction result than markets with three-part offers and locational marginal pricing and additional payments to guarantee recovery of start-up and no-load offer costs. Buyers may believe that a pay-as-bid rule will result in a lower market clearing price. This belief is sometimes based on the perception that in a competitive pay-as-bid auction, a unit with a $20/MWh marginal cost will always offer at that price, even if other units are clearing at a higher price.

This argument is incorrect. As long as there is sufficient market price transparency, then a supplier will always seek to obtain the price that clears the market to maximize profit. Hence, in a pay-as-bid market, the unit with a $20/MWh marginal cost will raise its offer price to its estimate of the price offer of the most expensive unit cleared in the same auction time period. This incentive raises the following concerns with pay-as-bid auctions. First, the need ~~by~~ sellers to estimate the hourly market clearing price could lead to many incorrect guesses, which even in a competitive market will lead to inefficiency (e.g., when a lower-cost unit overestimates the price, letting a higher-cost unit clear the market). Second, from the perspective of the market operator, efficient control of the power system will become less transparent as all offers for a particular hour will cluster around the market clearing price.

There is the argument that a low-price unit will be risk-averse in a pay-as-bid auction and rather than attempt to clear at the estimated market price will persistently shave its offer to ensure that it is scheduled. That is, a $20/MWh unit will offer at $50/MWh rather than at $60/MWh, which is the price of the most expensive unit chosen. Such risk aversion would lower the price to buyers. However, over time, it will also send the wrong price signal for investment, resulting in higher-cost plants being built. Hence, a short-term lowering of market prices may ~~shift to~~ an increase in prices over the long term (Cramton
AU3 and Stoft, 2006).

Finally, because all offers will likely cluster around the estimate of the market-clearing price in each auction period, if there is the need for *ex ante* or *ex post* market power monitoring and mitigation, the regulator will have more difficulty reconstructing which units have been raising their offers above marginal cost in an attempt to manipulate price levels. This is discussed later in this section and in Section 5.9.

For all these reasons, US regulators have preferred the uniform clearing price rule to the pay-as-bid rule. However, in at least one other country (England and Wales), regulators have actually redesigned the spot market to move away from uniform pricing rules, adopting pay-as-bid pricing.

5.2.4. Sequencing of markets and reliability functions

As noted earlier, the basic design choices for real-time electricity auctions are highly constrained by the characteristics of power system operations. Reliability requirements, timelines for operational decisions in the hours prior to real-time, and representation of generation and transmission capacity constraints all further shape the forward market designs by adding mathematical constraints to the auction solution and requiring specific market and "out-of-market" procedures. Over the decade of ISO auction market development, two major questions stand out in this regard:

- If there are forward energy or ancillary service auctions conducted by an ISO or any entity able to offer into an ISO market, how far forward in time should system reliability and operational constraints be applied? Put another way, how should the interface between forward and real-time markets be designed?
- How should the physical constraints of power system operations be reflected in the ISO day-ahead and real-time auction pricing rules?

Both of these questions were central in the market design discussions that took place in the United States in the mid-1990s (see, e.g., Stoft, 2002; Wilson, 2002; O'Neill et al., 2006). With regard to the first question, an initial debate placed those who believed that system operations could be a primarily real-time function, with forward markets operating separately under their own rules until just a few hours or less before real-time, against those who argued for a full integration of market and system operations. The second question concerned issues such as the choice of zonal pricing versus locational marginal pricing in transmission usage pricing and whether supply offers into the auctions should give generators the choice to represent details of their short-term marginal costs, such as start-up costs. The FERC standard market design proposal (FERC, 2002) sought to settle as many of these design debates as possible. As it proposed, and is now done in most of the ISO markets, the sequence of short-term ~~(i.e., daily)~~ auction markets and reliability actions incorporates all relevant reliability procedures and physical constraints and provides sufficient offer and bid detail to provide an efficient auction result. As the remainder of the chapter will explain, this sequence has the following structure:

- First, the ISO undertakes a series of pre-day-ahead procedures and out-of-market actions to account for changes in system conditions as well as demand (load) forecasting and scheduling of generators with longer than one-day start-up or shutdown requirements.
- Second, the ISO operates a *day-ahead market* that includes an auction with "security-constrained unit commitment," i.e., consideration of the full set of known transmission security and generation unit constraints, within the limitations of the auction optimization.
- Third, the ISO takes several types of additional actions after the day-ahead market clears to ensure reliability prior to real-time. Most notably, all ISOs undertake variations on what can be called *reliability unit commitment*, commitments of additional generation to meet forecasts of actual load if such forecasts are different from the amount of demand that is cleared day-ahead. In addition, the ISO collects data on generation cleared day-ahead to determine changes in actual availability.
- Finally, in the *real-time market*, as discussed earlier, the ISO operates the power system through a security constrained economic dispatch (complemented by unit commitments) using offers and bids to determine auction market prices and supported also by physical dispatch instructions that may ~~be~~ differ~~ent~~ from the market result for reliability reasons.

5.2.5. Market power monitoring and mitigation

Market power in the electricity auction markets is defined as the ability of a buyer or seller to significantly and sustainably alter the market price from the competitive price. Most economists interpret the competitive price as the market price that results when sellers are willing to offer power at their marginal opportunity cost and buyers bid for power at their true willingness to pay. The general economic principle is that the larger the number of sellers and buyers, the more likely that market prices and quantities will be competitive. This principle is measured through indices of market concentration or by market price simulations.[7]

Since there are currently few price-responsive buyers in electricity spot markets, the auction price is usually set by sellers (except when administrative scarcity pricing is enforced). Hence, the competitive price is typically estimated or simulated based on known production costs of the marginal unit delivering to a location, primarily fuel costs, and, if possible, adjusted to account for short-term fixed costs (such as start-up) and inter-temporal opportunity costs (for limited energy plants, such as hydro or emissions-constrained facilities). When the market price is above this competitive price, which it usually is, then some degree of supplier market power is being exerted. In economic terms, consumer surplus is being transferred to producers and the total producer surplus is being shifted among suppliers. The task of the regulator is to determine whether and how to manage such market power such that a reasonable approximation of competitive market prices and quantities prevails. The legal and regulatory methods for doing so vary between countries and supra-national organizations, such as the European Union. In the United States, FERC has the statutory obligation and authority to mitigate the market power of sellers and buyers in the wholesale electricity markets under its jurisdiction.[8]

Although there have been occasional concerns with buyer market power, the primary regulatory concern in the United States has been with seller market power.[9] In general, there are essentially four types of measures through which seller market power in an ISO auction market (or other electricity markets) could be mitigated. First, *ex ante* structural measures can be taken to enhance the competitiveness of the auction market prior to the market start. For example, if regulators find that the market is too concentrated, ownership of generation assets can be restructured, through divestiture, to diminish the market power of sellers. In the United States, FERC has not required any utility to divest generation prior to selling wholesale power; instead, FERC can selectively prevent suppliers with excessive market power from selling wholesale power at market prices by approving sales only at regulated cost-based rates.[10] In practice, this procedure has not

[7] Most ISOs now include such measurements of market power in their annual state of the market reports.

[8] In the United States, the Federal Power Act requires that generation and transmission market prices are "just and reasonable." This standard has been interpreted by the courts as giving FERC the authority to monitor and mitigate market power in the electricity markets. See discussion in O'Neill et al. (2006).

[9] With respect to buyer market power, a notable design decision was taken in the first phase of the California wholesale market design to require the incumbent vertically integrated utilities to purchase all their wholesale power through the daily markets in part to suppress their buyer market power. See, e.g., discussion in Sweeney (2006), p. 338.

[10] FERC screens prospective sellers, individually or collectively (e.g., as market participants in an ISO), for generation market power using various measures of market share and market concentration. Only sellers that pass the screens are automatically granted the right to sell at market prices (called "market-based rates"); those that fail must either prove their lack of market power using additional information or sell at a regulated cost-based rate or some other negotiated rate. See, e.g., Helman (2006).

been used to screen individual market participants in ISO markets. Rather, sellers are allowed to participate in the ISO auctions subject to the ISO's market power mitigation rules.

The ISO market power rules are predicated on the assumption that while the market is structurally competitive in the aggregate, some sellers have market power on an intermittent basis, such as when transmission constraints subdivide the market or during peak hours when most or all generators must run. This has led to a second type of market power mitigation: the application of *ex ante* behavioral measures. These usually take the form of offer caps that limit the price that a supplier can offer into the auction. The ISO auction markets have relied primarily on such offer caps, as discussed in Section 5.8. These caps can apply at all times to all bidders (as the $1000 offer caps do in the eastern US ISOs), or they can be selectively imposed if there is evidence that offer prices are raised significantly above competitive levels and also would affect market prices.

Third, the regulator and the ISOs can apply *ex post* measures, such as penalties for certain types of behavior or refunds to buyers if the auction market prices are found to be in excess of competitive levels, within some margin of error. FERC has the legal authority to apply such *ex post* measures, as it did, e.g., in requiring refunds following the California market crisis (see, e.g., Sweeney, 2006). It is a policy decision how to strike a balance between *ex post* and *ex ante* measures. In general, *ex ante* measures provide for ~~greater~~ market ~~certainty~~ than *ex post* measures. For example, offer caps can be seamlessly integrated into the auction pricing rules, as discussed in Section 5.8.

Finally, there are auction designs that mitigate market power without the need for structural or behavioral measures. These are called Vickrey–Clarke–Grove mechanisms, and they provide incentives for truthful bidding and, thus, efficient operation (Hobbs et al., 2000). Such an auction design could in theory be applied to electricity markets, but poses a number of practical issues. Perhaps the most serious issue is that such a mechanism would not be revenue neutral; in general, the auctioneer would pay more to suppliers (in a sense, to pay off their market power) than they would receive from market buyers. Hence, the preference of ISO market designers has been to retain the auction design with sealed bids and uniform price clearing, but apply various methods of *ex ante* and *ex post* market power mitigation.

5.2.6. Scarcity pricing

The advent of supply offer caps for purposes of market power mitigation and the lack of demand bids has created a new design issue for the electricity auctions, namely that spot market prices cannot always reach levels sufficient to reflect the true scarcity of supply or the true willingness-to-pay of demand. In markets without such economic regulation or market incompleteness, there is no distinct feature called "scarcity pricing." Rather, prices rise and fall largely in response to the supply–demand balance. Under scarcity conditions, supply is short and prices are ~~abnormally~~ high relative to historical norms, so consumers voluntarily or involuntarily ~~drastically~~ reduce their consumption of the commodity in question. In the US electricity markets, empirical evidence emerged that some existing generators in ISO markets were not revenue sufficient due to offer caps (and were seeking to return to regulated status under types of cost-based contracts) and that investment in both generation and demand-response capability was not adequate ~~despite the existence of capacity markets~~. Largely because the electricity markets are not considered sufficiently competitive in all market conditions to remove or greatly relax the offer caps, FERC and market operators have turned recently to administrative scarcity pricing as a means to complete the market design.[11]

As of this writing, various scarcity pricing rules are being implemented, and some proposed, in the ISO markets. The first design issue is to define a scarcity condition in the market. In the current electricity markets, with largely price inelastic short-term demand, scarcity is clearly indicated when demand is either voluntarily or involuntarily curtailed. From a system operational perspective, operating reserve shortages are the first indicator to the market operator that demand curtailments are more likely. Hence, most ISOs are using, or considering using, operating reserve shortages as a trigger for scarcity pricing. The second design issue is to determine what the scarcity price should be and how it should be set. The general proposed approach is to raise the price above the highest offer cap incrementally, corresponding to the degree of reserve shortage or demand curtailment. ~~Hence,~~ one approach is to create a "demand curve" that begins to automatically set the price as soon as reserves are short.[12]

Join these 2 paragraphs

The market price is determined by the intersection of the available reserve quantity and the demand curve, with the highest price being reached when reserves are at a minimum allowable level. A simpler method is to raise the price to a single pricing point when reserves are short~~. That is,~~ to create an adder to the energy market price. Most analysts would prefer that the scarcity pricing curve ~~is~~ related to a measure of the value of lost load (VOLL) (see, e.g., Stoft, 2002, 154–64). However, there is no consensus on how to determine VOLL exactly, and high VOLL prices may be politically unacceptable. Hence, as will be discussed further below, while different ISOs have reached their own conclusions, there is as yet no consensus on the parameterization of scarcity pricing curves.

With this background on how auction design decisions have been made and how regulatory, technological and administrative factors have ~~required~~ modifications to the auction designs and pricing rules, the chapter turns next to the details of the spot electricity auctions, beginning with the day-ahead market.

5.3. The Day-Ahead Market

The day-ahead market begins the sequence of short-term market-driven and operational procedures that lead to the daily efficient and reliable functioning of the regional power system under ISO control. The day-ahead market encompasses a day-ahead auction market for energy as well as regulation and operating reserves that follows the basic auction format described in Section 5.2, along with many other details that will be described here. The day-ahead market also includes non-price schedules that are included as constraints in the auction market clearing. These schedules include generation self-schedules, bilateral schedules, and import and export schedules.

As currently designed, the day-ahead market is best described as a forward market subject to all the physical and reliability power system constraints that are known at the time to affect the next-day (real-time) dispatch. It is a forward market because sales or purchases cleared at the day-ahead price that are not subsequently converted into a physical position in real-time must be "sold back" or "bought back" at the real-time market

[11] The initial expectation in most ISO markets was that the $1000/MWh offer cap would ~~be sufficient to~~ allow suppliers to set a sufficient scarcity price using their offers (see, e.g., FERC, 2002).

[12] Under scarcity pricing ~~supply offers are suspended and~~ the price is set by the administratively determined ~~method~~.

price.[13] This design principle generally applies to any generation product sold in the day-ahead market, whether energy, regulation, or operating reserve (although as discussed below, in practice ISOs may not fully implement this principle for reserves). It also applies to energy purchased in the day-ahead market, but not to regulation or reserves (which the ISO procures on behalf of buyers). This sequence of financial settlement generally creates an inducement to transact in the day-ahead market based on a forecast of the next-day supply–demand balance.

The "physical" aspect of the day-ahead market is that the auction financial offers, along with any submitted schedules, are subject to a number of physical constraints on market clearing, including generator constraints, such as ramp rates and minimum and maximum output levels, and transmission network constraints. With the inclusion of offers to start-up generation units, this is called "security-constrained unit commitment." The initial objective of market designers was that security-constrained unit commitment, in concert with the financial incentives noted above, would shape the day-ahead schedule into a reasonable approximation of a feasible real-time dispatch. This would give the ISO power system operators time to evaluate system conditions in the spot markets, under which suppliers are no longer necessarily dedicated to particular buyers, before the operating day began. However, with the introduction of "virtual" sellers and buyers (defined in the next section), which can displace physical offers and bids, the day-ahead schedule has become less physical and more financial. To address this, ISOs introduced additional reliability unit commitments following the day-ahead market, as described in Section 5.4. Hence, the day-ahead market, taking into account all relevant system constraints, in concert with the reliability unit commitment allows the ISO to reduce scheduling uncertainty about the next-day dispatch.

5.3.1. Market procedures

The day-ahead market trading rules are fairly consistent across the ISOs. Offers and bids in the day-ahead market, along with non-price schedules, are due typically by midday of the prior day, although as shown in Table 5.3 some ISO markets open and close earlier T3
in the day. Offers and bids for energy and ancillary services are divided into price offer components and physical parameters. These are summarized for two ISO markets, PJM and New York ISO, in Table 5.7. As noted already, in each day-ahead ISO market there are three price components for energy suppliers, which also support the provision of regulation and reserves: start-up ($), no-load ($/MWh), and energy, sometimes called "incremental" energy ($/MWh). The energy offer represents the seller's minimum dollar value that it is willing to accept to supply energy. A negative energy offer represents the seller's maximum willingness to pay to produce power at a particular level of output (usually the minimum physical load level of the generation unit). The markets for regulation and operating reserve may include an additional offer component ($/MW) to represent costs associated with providing those services. The remaining generation offer components are physical parameters, which include upper and lower operating limits, both under normal and emergency conditions, ramp rates, minimum run times, maximum starts per day, and other parameters. Demand bids for energy may also have multiple components. In general, the core feature of demand bid is a MW block with a $/MWh reflecting the buyer's

[13] In US regulatory jargon, this is sometimes called a "two-settlement system," referring to the day-ahead (first) settlement and the real-time (second) settlement.

Table 5.3. Scheduling, ricing, and settlement timelines for day-ahead and real-time energy in the US ISOs

18:00

by unit

	PJM	New York	New England	MISO
Day-ahead market offer period closes (prior day)	12:00	05:00	12:00	11:00
Day-ahead market results posted (prior day)	16:00	11:00	16:00	
Reliability unit commitment opens for offers	~~18:00 (prior day)~~	Integrated with day-ahead market		17:00
Reliability unit commitment offer period closes		05:00		18:00
Reliability unit commitment results posted		11:00		20:00
Real-time market opens for offers	16:00 (prior day)	~~18 months prior~~	16:00 (prior day)	
Real-time market offer period closes	18:00 (prior day)	75 minutes prior to the (dispatch) hour	18:00 (prior day)	
Real-time market preliminary results posted		30 minutes prior to the (dispatch) hour		During dispatch hour (integrated 5 minute LMPs)
Real-time market final settlement results		Within 2 business days after Operating Day (but up to 5 business days in exceptional circumstances)		

16:00

During dispatch hour (integrated 5 minute LMPs)

30 minutes prior to the (dispatch) hour

Sources: ISO and RTO technical manuals (see references).

REMOVE LAST ROW

maximum willingness to pay to consume power. Demand bids may also submit fixed cost components, such as an additional cost to shut down a particular piece of equipment. As Table 5.7 illustrates, ISO vary in the periodicity of day-ahead offers and bids and the frequency with which they can be changed. The implications of these differences will be discussed below.

Beginning in 2000, the eastern US ISOs began to introduce virtual supply offers and demand bids into the day-ahead markets. These are purely financial positions in the forward market that will not be converted into a physical position and must be re-settled at the real-time market price. An accepted virtual supply offer that "sells" energy at a day-ahead LMP has to "buy back" that energy at the real-time price at the same location. Similarly, an accepted virtual demand bid that "buys" energy day-ahead has to "sell" it back in real-time. There are a number of purposes for such virtual transactions. A common use by purely financial entities (i.e., those that have no physical positions) is to arbitrage the price spread between the day-ahead and real-time markets. Hence, a virtual seller seeks to sell day-ahead energy at a higher price than it will have to buy the energy back in real-time; a virtual buyer has the opposite objective (see the numerical examples in Sections 5.3 and 5.5). With entry into the virtual trading market, competition between virtual traders

Els UK Job: CDI ch05-i047172 25-10-2007 7:33p.m. Page:199 Trim:165×240MM Float:Top/Bot TS: Integra, India

for arbitrage rents causes greater price convergence between day-ahead and real-time. There are other uses for virtual transactions. For example, if an entity holds financial transmission rights, it can use such virtual positions to shift settlement of the transmission rights from day-ahead to real-time, if that is seen as financially advantageous.[14]

5.3.2. Market pricing and settlement

The day-ahead auctions basically follow the uniform-price pricing rule discussed in Section 5.2.3 for energy, but with additional features to pay other offer components and with many other pricing rules. Prices are determined as follows: The auction is conducted by inserting the supply offers and demand bids into a dynamic optimization program that minimizes the cost of meeting demand (or maximizes social welfare, measured as the sum of consumer and producer surplus) for energy and ancillary services in each hour of the operating day subject to generation and transmission constraints. The generator constraints are represented in the auction model both through discrete, or integer, variables, such as the start-up decision, and continuous variables, such as the generator energy offer (represented as a stepwise or piecewise linear function over the range of output). The day-ahead auction result is an hourly schedule of LMPs and non-binding dispatch instructions for each generator, indicating its output level and respecting its start-up costs, energy offers, ramp rates, and other constraints on output, such as minimum and maximum operating limits. ~~In electricity parlance, this is called "unit commitment" (see, e.g., Hobbs et al., 2001).~~

The market price for energy is calculated by the auction algorithm as the shadow price associated with the energy balance constraint for each node (typically a bus on the network) in the high-voltage transmission network (see Appendix 5A). That is, LMPs can be calculated for each network location where energy is injected or withdrawn and also for transshipment nodes. ISOs will typically limit the number of LMPs calculated to nodes that have commercial purposes. The LMP is a composite of the accepted offer prices of all generators that would supply the next, or incremental, MW at that location. ISOs typically disaggregate LMPs into three components: energy, losses, and congestion (Schweppe et al., 1988). For computational reasons discussed in Section 5.7, the energy component is the same for all buses within an ISO's network, while the loss and congestion components differ if there are losses and congestion.

The ISO will generally collect surplus revenues when there are losses and congestion. Because the ISO is revenue neutral, these surpluses are returned to market participants through financial transmission rights and loss refunds. One reason for disaggregating the LMP is that the financial transmission rights issued by ISOs typically only cover differences in the congestion component, and not the loss component. Hence, the ISOs

---

[14] A financial transmission right as currently specified in the ISO markets is a contract to receive the difference in price between two locations on the transmission network for a particular quantity (MW) at a particular time. Financial transmission rights are cashed out in the day-ahead market. An entity holding a financial transmission right can take an equivalent virtual position to create the exact opposite of the locations and MW specified in the right for settlement purposes. That is, it would bid to buy equivalent MW at the location where it was to withdraw the MW specified in the transmission right and offer to sell at the location where it was to inject the MW specified in the transmission right. This would exactly cancel its financial transmission right position day-ahead, but would require it to cash out the equivalent position in real-time due to the virtuals. So the result would be that it buys back at the injection location and sells back at the withdrawal location at their respective real-time prices.

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T.W:30PC open recto 1 Color 49 Lines

need to calculate how much surplus is generated by each component. There are various ways to calculate these components and the corresponding auction surpluses, which will be discussed in the numerical example that begins in this section as well as in Section 5.7.

While the day-ahead LMPs will always provide a generator accepted through the auction with sufficient revenues to cover its energy offer price over the hours that it operates, those prices may not cover its start-up and no-load costs. The numerical example that begins in this section illustrates this possible outcome. The ISO ~~guarantees that all unrecovered~~ offer costs will be recovered through a revenue sufficiency guarantee. This is an additional payment that takes place at the end of the day-ahead market, the real-time market, or both markets. This payment is discussed in more detail in Section 5.6. ~~This additional payment adds an element of "pay-as-bid" to the auction, although not one that affects the day-ahead pricing of energy.~~

In general, all generators that provide a price offer for energy and all price-responsive demand bids that are "dispatchable" and all virtual offers and bids are eligible to set day-ahead LMPs. A dispatchable generator or demand resource is one that the ISO can dispatch up or down based on its offered supply function. In addition, some ISOs let congestion bids (i.e., offers for a maximum price difference between two nodes) set locational energy prices. However, dispatchable generators in some situations are not allowed to set the price. There is usually a straightforward operational and/or economic incentive reason for such a restriction. ~~First,~~ generators that are scheduled at their minimum load level for some hours are not eligible to set the locational price~~, but are required to be price takers (and are also typically eligible for the revenue sufficiency guarantee)~~. The reason for this rule has to do with incentives: if the generator could set the price in the hours scheduled for minimum output, it may try to produce energy to increase its revenues. This would undermine the ISO's attempt to establish an efficient schedule that respects inter-temporal constraints. ~~Second, generators whose offers have been adjusted for purposes of market power mitigation (as discussed in Section 5.8) are also generally not allowed to set the price. Generators that are not dispatchable are typically required to be price takers.~~

In both the day-ahead and the real-time markets, most ISOs have instituted various *ex post* aggregations of LMPs for settlement purposes. One such aggregation is called "hub" pricing, in which a location-weighted price is calculated for a set of nodes where spot energy is sourced. Another such aggregation is "zonal" locational pricing for demand, in which a load-weighted average price is calculated for the territory of particular utilities that have retail customers. In some cases, such zonal pricing is used to settle the purchases of multiple utilities, thus embodying some level of cross-subsidy. Some of these zonal pricing subsidies reflect regional agreements among utilities and state regulators. However, such zonal pricing also inhibits the development of demand response, as the actual nodal price is not known by buyers and higher prices will be hidden by the averaging process.

Transmission usage charges apply to any market participant that has ~~both injections and~~ withdrawals in the market. This pertains both to spot transactions and to non-price schedules. With respect to the latter, a buyer with a bilateral contract or a utility that remains vertically integrated and desires to operate its own generators has no requirement to purchase energy through the day-ahead (or real-time) market. However, there is a requirement by such schedulers to pay for marginal transmission usage, as measured by differences in spot LMPs between points of injection and withdrawal. As discussed in more detail in Section 5.7, the congestion component of such usage charges for any particular MW schedule are hedged by financial transmission rights between those points for the equivalent MW schedule, while any surplus marginal loss charge payments are refunded by the ISO on some basis independent of particular schedules.

5.3.3. Markets for regulation and reserves

As noted in Section 5.2.3, some day-ahead markets now incorporate co-optimized regulation and operating reserves markets, although others have only real-time markets for these products. Details on this procedure are provided in Section 5.3.7.[15] The design principles for day-ahead auctions of these products generally follow the sequence of financial settlement obligations described earlier. Eligible generators and responsive demand make offers for these services into the day-ahead market. ISOs also allow "self-supply" of regulation and operating reserves, whether from a utility's own resources or via contract. In general, to ensure reliability, the ISOs require that these schedules are offered through the auction at a zero price. Unlike energy, buyers do not make specific bids for these products into these markets, which are procured on their behalf by the ISO. The quantities that clear in the day-ahead market are then transferred to the real-time market and deviations are priced at real-time prices.

5.3.4. Scarcity pricing

As discussed in Section 5.2.6, ISOs are experimenting with alternative types of scarcity pricing that set market prices during operating reserve shortages. In general, the ~~"~~demand curves~~"~~ or price adders/caps used to determine scarcity pricing in the day-ahead market should be the same as those used in the real-time market, to avoid providing incentives to market participants to alter behavior in one market so as to affect the prices in the other. However, there is a practical difference in the triggering mechanism. Day-ahead, an administrative scarcity price will be determined instantaneously as part of a day-ahead market solution. In real-time, as discussed below, system operators may be taking non-market steps to maintain reliability, and hence the declaration of an operating reserve shortage will be subject to manual decision-making within the operating hour. Some specific scarcity pricing rules are discussed below.

5.3.5. Congestion management

Because congestion management is a slightly different procedure ~~than~~ day-ahead ~~and~~ real-time, the day-ahead procedure is described briefly here. Essentially, congestion in ISOs with LMP is managed day-ahead "implicitly" through the auction market optimization. Put simply, the day-ahead market result is a schedule for each hour of the day that includes the effects of congestion and in which the congestion "charge" between any two nodes or buses on the system is the difference between the LMPs at those locations. In contrast, in real-time, congestion management is an ongoing process that requires the system operator to make operational adjustments from minute to minute and which may involve non-market decisions. That procedure will be discussed in Section 5.6.

5.3.6. Numerical example

This section begins a numerical example of market pricing and settlements that will continue through several sections of the chapter to illustrate the different stages of the ISO

[15] Co-optimization is less cumbersome day-ahead than in real time, because the real-time procedure requires constant updating of generator set points and available regulation or reserve capacity. Real-time co-optimization is described in Section 5.5.7.

energy auctions. The example is developed in the simplest fashion possible to demonstrate realistic characteristics of an auction market with locational marginal pricing, while allowing the reader to replicate and track auction results. The mathematical details of this example are provided in Appendix 5A.

The example assumes a transmission network with three transmission lines connecting F1 three ~~electrical~~ buses, or nodes, as shown in Fig. 5.1. A bus is represented by a thick vertical line while a transmission line is the thin line or lines connecting the buses. The network has characteristics that allow for the calculation of both the marginal congestion and the marginal loss components of LMPs. The transmission line connecting buses 1 and 2 has a capacity limit of 350 MW in both directions (1→2, 2→1); the other lines do not have capacity limits that will affect the examples. To simplify the presentation of power flows on this network, a DC approximation of the AC load flow is used, in a version with quadratic line losses (Schweppe et al., 1988). This means that for any 1 MW injected at one bus and withdrawn at another bus, the percentage of that 1 MW that is withdrawn is a decreasing, non-linear (quadratic) function of the total MW flowing on the line (the line "loadings"). As noted, all the US ISOs calculate the loss component of LMPs, so while adding losses to the example makes it somewhat less intuitive, it is reflective of how the actual prices are calculated. The power flow equations that are being used in these examples are shown in Appendix 5A.

All the demand in this example is located at bus 3; this simplifies the presentation, as all power will flow to this location in each example. The 24 hours of the day-ahead market are compressed here into three demand blocks: an off-peak demand of 950 MW, an intermediate demand of 1300 MW, and a peak demand of 1600 MW. To further simplify, these demands are price-inelastic. Virtual demand bids are not considered, but would be represented as price-sensitive demand blocks (as would any price-sensitive physical demand).

T4 There are five potential suppliers, whose parameters are found in Table 5.4. The location of the suppliers is shown in Fig. 5.1; physical generators are represented as the circles connected by a solid line to each bus, while the one virtual supplier is represented as a circle connected by a dashed line. A cheap "base-load" generator "A" is located at bus 1. The offer price ~~on~~ this generator is \$15.00/MWh and its capacity is 1500 MW. A more expensive "intermediate" load generator "B" is located at bus 2. The offer price ~~on~~ this generator is \$20.00/MWh and its capacity is 250 MW. Both of these generators have start-up costs. The base-load generator has a very high start-up cost but because it is assumed to be operating for most hours of the year, and hence is already running in hour 1 of the day-ahead market, the auction algorithm does not need to consider its start-up cost.

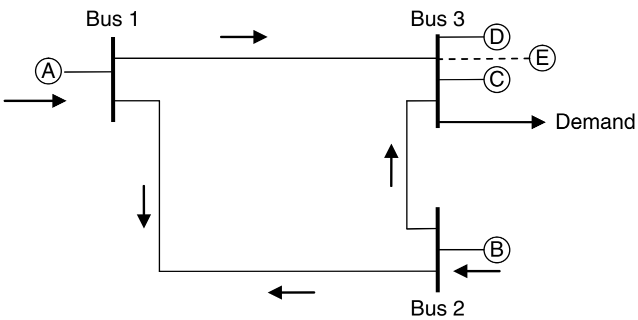

Fig. 5.1. Three bus, five supplier electrical network.

Table 5.4. Supply parameters and offer prices

	Bus	Capacity (MW)	Energy offer ($/MWh)	Start-up price ($)
Generator A	1	1500	15	Not applicable
Generator B	2	250	20	1000
Generator C	3	200	40	2000
Generator D	3	300	50	100
Virtual E	3	200	41	0

Table 5.5. Results of day-ahead market simulations

Scenario	Demand (MWh)	Supply (MWh)					Price at bus ($/MWh)		
		A	B	C	D	E	1	2	3
Off-peak	950	1018					15.00	16.13	17.31
Intermediate	1300	1227	190				15.00	20.00	20.36
Peak	1600	1288	250			197	15.00	54.49	41.00

The intermediate generator has a start-up cost of $1000. At node 3, there are two peaking generators, "C" and "D," with similar capacity, but different offer prices, as well as one virtual offer, "E." Peaking unit C has an energy offer price of $40.00/MWh and a start-up price of $2000; peaking unit D has a higher energy offer price of $50.00/MWh, but a lower start-up price of $100. The virtual offer E is priced at $41.00/MWh to exploit the price gap between the peaking units. Virtual supply offers could also exploit other jumps in the supply function in this example.

The results of each demand scenario are summarized in Table 5.5 and in the Figs 5.2a–c. Beginning with the off-peak demand scenario depicted in Fig. 5.2a, the base-load generator at node 1 can supply all the power needed to meet demand of 950 MWh. The transmission capacity constraint on line 1 ↔ 2 does not bind, i.e., does not create congestion. However, the quadratic line losses require the base-load generator to inject 1018 MWh to meet the load (meaning that approximately 68 MWh are lost due to line losses) and the losses also establish differences in the LMPs. The prices at buses 1, 2, and 3 are $15.00/MWh, $16.13/MWh, and $17.31/MWh, respectively.

At the LMPs for the off-peak scenario, the ISO auction charges demand $17.31/MWh × 950 MWh = $16 444. The ISO auction pays the base-load generator $15.00/MWh × 1018 MWh = $15 275. The difference between what the ISO collects from demand and what it owes to generators is an ~~auction~~ surplus, which in this case is $1169. In this scenario, because there is no congestion, the surplus is due to charging demand for marginal losses. Since the ISO is revenue-neutral, it must refund this surplus in some fashion, as discussed in Section 5.7.

In the intermediate demand scenario depicted in Fig. 5.2b, while the demand of 1300 MWh is lower than the base-load generator's capacity, the transmission capacity limit of 350 MW on line 1 ↔ 2 now creates congestion that prevents the base-load generator from meeting the demand on its own.[16] This requires dispatching generator B at bus 2 to

[16] The base-load generator congests line 1 → 2 when it injects around 1038 MWh.

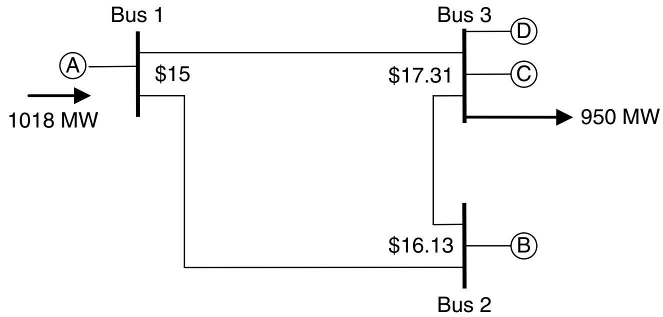

Fig. 5.2a. Off-peak scenario.

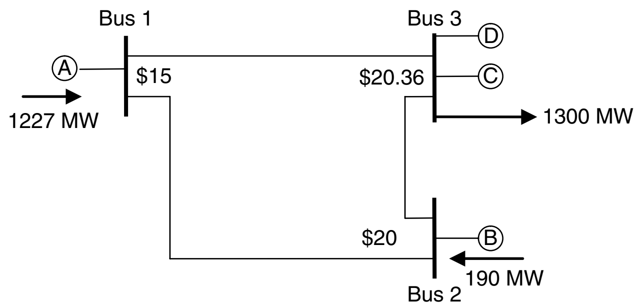

Fig. 5.2b. Intermediate scenario.

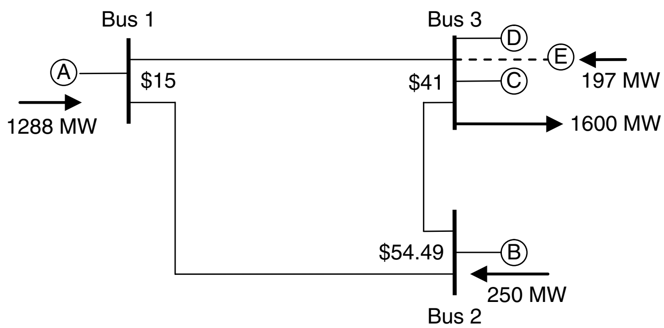

Fig. 5.2c. Peak scenario.

provide "counterflow" on line 1 $\leftrightarrow$ 2 that can resolve the congestion and also meet the demand at bus 3. In the day-ahead auction, as noted above, this congestion management happens simultaneously in the solution of the auction. The intermediate generator has a start-up cost of $1000 that is considered in the auction's decision to commit this generator. However, as noted above, this cost does not enter into the calculation of the LMPs. The prices at buses 1, 2, and 3 are $15.00/MWh, $20.00/MWh, and $20.36/MWh, respectively. Using bus 1 as the slack bus for the purpose of calculating LMP components, the energy component of the price is \$15.00/MWh, while the loss (congestion) components are \$0 (\$0), \$1.18 (\$4.82), and \$3.19 (\$1.17) per MWh, respectively at the three buses.

At these intermediate scenario prices, the ISO auction charges the demand at bus 3 \$20.36/MWh × 1300 MWh = \$26 468. The ISO auction pays the base-load generator \$15.00/MWh × 1227 MWh = \$18 411, and the intermediate generator \$20.00/MWh × 1[illegible] MWh = \$3793. The surplus collected by the ISO is now \$4264, which is due in this sce[illegible] both to marginal loss charges and to marginal congestion charges. Using the above [illegible]components, the congestion surplus portion of this total surplus is \$2163. This means [illegible]e loss component is \$2101, equaling the sum of the loss LMP components times the [illegible]thdrawals (\$3857) minus the energy LMP component times the net losses (\$1756). [illegible]wever, this division is arbitrary. If the LMP components were instead based on using [illegible]ad bus as the swing bus ~~(as in the CAISO, 2005 methodology)~~, the estimates of [illegible]rpluses results would have been different. ~~Then the energy component would be \$20.36 (the bus 3 price), and the loss components would have been −3.52, −2.20, and 0.00 \$/MWh at the three respective buses.[17] The resulting congestion components would also be negative, being −1.84, 1.84, and 0.00, respectively. The congestion surplus would then be calculated as (−1.84 × −1227 + 1.84 × −190) or \$1920. Subtracted from the total surplus of \$4264, this yields a loss surplus of \$2355. Thus, the loss and congestion surpluses based on bus 3 being the slack are each about 10% different from the values based on a bus 1 slack, above. Finally, if the congestion surplus was instead defined based on flowgate shadow prices, it would instead equal \$5.39 (the price for the congested flowgate between buses 1 and 2) times 350 MW (the corresponding flow), or \$1886; this would imply a loss surplus of \$4264 − \$1886 = \$2378. These values are close but not identical to the component surpluses resulting from using the distributed load slack (bus 3) as the slack bus.~~

THESE LINES HAVE BEEN DELETED BECAUSE THEY ARE REPEATED IN SECTION 5.7

In the peak demand scenario in Fig. 5.2c, both the base-load generator and the intermediate generator are operating at the highest output possible given the congestion on the system, but meeting the demand of 1600 MWh requires using one or more peaking units, which in this example are located close to the load at bus 3. Because generator A is still not operating at full output, it remains a marginal generator and the price at its bus remains at \$15.00. However, generator B is operating at full output and the price at its bus has risen above its offer price, to \$54.49/MWh. This price is the value (in terms of reduced operating cost of the other operating generators) of a hypothetical additional MW (or increment) of power injected at bus 2.

The ISO auction has the choice of three supply offers at bus 3, each with different energy and start-up prices. Each of these offers alone is sufficient to meet the demand. The auction unit commitment decision is to pick the generator offer that either (a) minimizes the total offer cost of both energy and start-up if demand is not price responsive or (b) maximizes social welfare defined as the difference between buyer surplus and seller surplus (see Appendix 5A). Mathematically, this is called the auction "objective value." To illustrate the commitment decision, the results of each choice, including the resulting LMPs, are summarized in Table 5.6. As can be seen, the minimal objective function value T6
is to dispatch the virtual supply offer. This offer has a higher price than generator C, but has no start-up cost. The start-up cost of generator C is high enough that its lower energy offer is not sufficient to yield a lower total cost of energy and start-up. Generator D has

[17] For example, a 1 MW load increment at bus 1 would be met by decreasing load at the "distributed load slack" (just bus 3) by 0.827 MW; thus, losses would be lowered by 0.173 MW, which at \$20.36/MWh is worth \$3.52.

Table 5.6. Comparison of unit commitment choices for peak demand scenario

	Commit Gen C	Commit Gen D	Commit Virtual E
Energy offer price ($/MWh)	40.00	50.00	41.00
Start-up offer price ($)	2000	100	0
Nodal price at bus 1 ($/MWh)	15.00	15.00	15.00
Nodal price at bus 2 ($/MWh)	52.80	69.75	54.50
Nodal price at bus 3 ($/MWh)	40.00	50.00	41.00
Total demand payments for Energy + start-up ($)	62.000	80.100	65.600
Total supply payments for Energy + start-up ($)	42.395	46.702	41.015
Auction objective function value ($)	34.196	34.265	32.393

a lower start-up cost than generator C, but its higher energy cost results in a higher total supply cost. Note that total payment to sellers is higher than their total offer costs because the LMPs can be higher than the generator offer price at a location. Put another way, some generators are "inframarginal": in this example, generator B at bus 2.

The virtual supply offer is priced in this example to displace Gen C. The expectation of the virtual supplier is that in real-time, Gen C will set the price, allowing the virtual to sell back its position at a lower price than what it has been paid in the day-ahead market. The example is continued in Section 5.5 to show the conditions under which the virtual supply offer makes positive revenues or faces a loss.

5.3.7. Comparison of PJM and New York ISO market rules

The general description of the auction markets given above is largely consistent across the US ISOs. For example, with respect to the day-ahead markets, all the ISOs operate auction markets for energy with security-constrained unit commitment which calculate hourly LMPs. However, there are many minor and several quite significant differences among the ISO market rules and procedures that affect market behavior and market prices. Beginning in this section and continuing in several subsequent sections, some of these differences between PJM and New York ISO will be examined, so as to further illustrate design choices and trade-offs. A thorough examination of the differences between these two markets, many of which stem from the preferences and expectations of the market designers and market participants in each region, is beyond the scope of this chapter. Readers are encouraged to turn to the ISO tariffs, technical manuals, and annual state of the market reports for further comparative analysis.[18]

5.3.7.1. General scheduling procedures and energy market rules

Turning first to the day-ahead energy markets, some of the basic differences between the auctions in PJM and New York ISO lie in the rules for specifying and submitting
T7 energy offers, the components of which are listed in Table 5.7. As this brief survey will suggest, there are some ways in which the PJM offer rules are more restrictive than those

[18] These are available on the ISO websites.

Table 5.7. Offer and bid components in the PJM and New York ISO short-term markets

	PJM		New York	
	Parameters	Variability	Parameters	Variability
A. Supply Offers and Demand Bids				
Generation Start-Up Price	$/hour	Six months	$/hour	hourly
Minimum Generation Energy Block and Price	MW, $/hour		MW, $/hour	hourly
Dispatchable Energy	Piecewise linear function: 10 pieces, $/MWh for the two points of piece, slope of line	Daily	Stepwise linear function: 11 steps, $/MWh, MW/Step	hourly
Demand Shutdown Price (fixed cost)	$/period	Daily		
Demand Bid	$/MW	Daily	$/MW	hourly
Regulation Capacity Availability	MW	Daily	MW	hourly
Regulation Price	$/MWh	Daily	$/MW	hourly
Spinning Reserve Price	$/MWh	Daily	$/MW	hourly, day-ahead only
10-Minute Non-Synchronized Reserve	n/a	n/a	$/MW	hourly, day-ahead only
30-Minute Operating Reserve	n/a	n/a	$/MW	hourly, day-ahead only

(*Continued*)

MW

MW/min

Add "Variability subject to PJM market rules" in this space (to take up three rows)

Table 5.7. (*Continued*)

	PJM		New York	
	Parameters	Variability	Parameters	Variability
B. Physical Generation Characteristics				
Dispatch Status			Whether ISO or self-committed	May vary
Start-Up Time	hours, min		hours, min	May vary per commitment period, day-ahead or real-time
Minimum Run Time	hours, min		hours, min	
Minimum Down Time	hours, min		hours, min	
Max. Start-Ups per Day			1–9	Static
Normal Upper Operating Limit			MW	May change over day
Emergency Upper Operating Limit	MW		MW	May change over day
Temperature-Based Operating Limits	MW plotted against temperature	Limited to Combustion Turbines		
Normal Response Rate			MW/min	May vary
Regulation Response Rate			MW/min	May vary
Regulation Maximum and Minimum	MW	Daily; can be increased but not decreased after day-ahead market closes		
Emergency Response Rate			MW/min	May vary
Reactive Power Capability			MW plotted against MVARs	Static
Physical Minimum Generation Limit			MW	Static

Note: Additional details are found in the PJM and NYISO tariffs. Static refers to offer components that remain relatively constant over the life of the offer, but can be changed.

in New York ISO, and other ways in which the opposite is true. One notable difference is in the ability to modify offers from hour to hour. PJM requires that a single energy offer, specified as a piecewise linear curve with up to 11 prices, be submitted for the full day (i.e., the same supply function for each hour), whereas New York, which specifies a stepwise linear function with up to 10 steps, allows offer prices to change from hour to hour. In addition, PJM restricts changes in start-up offers to once in every 6 months, whereas New York allows hourly changes. Hence, with respect to variability over time, the PJM offer rules are more restrictive than those in New York. For some generators that have the option to sell into neighboring markets, or potentially to make a bilateral sale within the ISO market, the hourly flexibility to change offers may reflect changing opportunity costs from hour to hour. In the real-time market, as discussed in Section 5.5, these differences in offer flexibility may have other implications for pricing.

Another reason for the differences in offer price flexibility is that the two markets have different market power mitigation rules. These rules will be discussed in more detail in Section 5.8. In short, in PJM, if a generation unit is not transmission constrained, it can generally submit any price offer up to \$1000/MWh without any consequence, while in New York, there are offer price screens that may result in an offer being mitigated to a reference price. Hence, in unconstrained areas, PJM appears to permit greater offer pricing flexibility than New York. On the other hand, for units that are in transmission constrained locations, PJM automatically mitigates offers to a pre-submitted marginal cost offer if they fail a market concentration screen (PJM, 2006b), while New York applies offer thresholds according to a formula that becomes progressively more restrictive with the number of congested hours, as summarized in Table 5.13.

One of the high-level differences between the market designs in New York and PJM is that the former sought to establish a price basis for as many market features as possible, while the latter retained several physical scheduling features that were more consistent, at least initially, with prior utility practice. For example, PJM allows participants to submit non-price energy schedules, whether bilateral schedules or self-schedules. That is, a generator can simply schedule itself to run in PJM, regardless of market price, as long as the schedule does not cause a reliability concern. Similarly, PJM allows for physical reservations for import schedules on its boundary. In contrast, New York has required that all schedules, including imports, have price offers. A supplier can attempt to guarantee that a generator runs, or that an import is accepted, by reducing its offer prices: the start-up price offer and no-load price offer can be reduced to \$0, while the energy price offer can be reduced to \$0/MWh or a negative price. However, if a negative price offer is accepted and sets the price, then the generator will be paying to run (NYISO, 2004).

In the case of imports, the price offer requirement was an issue for many years due to scheduling difficulties that it created due to software problems; other problems for imports stemmed from lack of co-ordination between the ISO markets (Potomac Economics, 2003, pp. 68–74). The upshot is that scheduling only through financial offers and market prices can occasionally create complications for market functioning, while physical scheduling may result in economic inefficiency. Finding the appropriate balance requires continuing refinement of market rules, software, and inter-ISO co-ordination.

5.3.7.2. Markets for regulation and reserves

New York ISO began market operations in 1999 with auctions for energy, regulation, and three operating reserves. In contrast, PJM operated only an energy market for several years, procuring ancillary services on a cost basis. PJM introduced a market for regulation

in 2000, followed by a market for spinning reserve, called synchronous reserve, in 2002 (see discussion in PJM, 2003). However, the design of its ancillary service markets has been
AU4 different in New York, both in the sequence of the markets and in their pricing rules. First, the PJM markets for regulation and synchronized reserve clear in real-time (although price offers have to be submitted on the prior day) and so will be discussed in Section 5.5.6 in the context of the PJM real-time market. New York ISO has both day-ahead and real-time markets for regulation and operating reserves that are simultaneously co-optimized with the energy markets. The rules for these markets are quite complicated and not all the details are covered here (see in particular NYISO, 2006).

In New York, day-ahead offers for regulation and reserves are subject to the same submission deadlines as energy. Regulation offers must include a regulation response rate (MW/min), which can be no less than 1 MW/min, and a regulation availability price, in $/MW. The three types of operating reserves are 10-minute spinning reserves, 10-minute non-synchronized (non-spinning) reserves, and 30-minute reserves, which can be provided from spinning or non-synchronized units. As shown in Table 5.7, suppliers of operating reserves can submit an "availability" offer ($/MW) for each hour of the day-ahead market and must also provide an emergency response rate that will be used during reserve pickup events. The remainder of the offer consists of physical parameters.

The maximum capability for different reserves is a function primarily of the energy output of the unit in relation to its upper operating limit and the unit's ramp rate.[19] New York currently has three locations (zones) for the operating reserve markets: the western zone, which is defined as west of the Central–East transmission path; the eastern zone, which is east of the Central–East transmission path, excluding Long Island; and Long Island. Each of these locations has a reserve requirement, with higher requirements in the east where most of the load is concentrated (NYISO, 2006).[20] The pricing of operating reserves is done by zone: there is a different price for each zone described above. Within each zone, the price of each type of reserve under normal operating conditions is done in a fashion that reflects the hierarchical ranking and substitution properties.[21] In addition, when operating reserve capacity is less than reliability requirements, prices are set by demand curves described next.

[19] In New York, the rules are that spinning reserve MW are calculated as the unit's emergency response rate multiplied by 10 (minutes); 10-minute and 30-minute non-synchronized reserve MW are the unit's upper operating limit (normal or emergency), and for synchronized 30-minute reserves, the emergency response rate multiplied by 20 (minutes), which gives the available reserves above the unit's 10-minute capability.

[20] Currently, the most severe contingency in New York state is rated at 1200 MW and the total 10-minute reserve requirement for the system is set at this quantity. Half of this total is required to be 10-minute spinning reserve, with half of the total spinning reserve purchased in the eastern location; the remainder can be either spinning or non-synchronized 10-minute reserves and must all be purchased on the eastern location. The 30-minute reserve is set at 150% of the most severe contingency, equal to 1800 MW, of which 1200 MW is purchased in the eastern location and 600 MW in the western location.

[21] Each surplus higher quality reserve offer can fulfill the quantity requirement of a lower quality reserve. The price of 10-minute spinning reserve is equal to the shadow price on the constraint for that reserve in the auction algorithm plus the shadow price for 10-minute non-synchronized reserve plus the shadow price for 30-minute reserve. Similarly, the shadow price for 10-minute non-synchronized reserve is the shadow price for that reserve plus the shadow price for 30-minute reserve. Finally, the price for 30-minute reserve is the shadow price on its constraint.

5.3.7.3. Scarcity pricing

When supply is tight, reflecting scarcity, the two ISOs have different procedures for scarcity pricing. PJM essentially ~~requires~~ suppliers to voluntarily raise offer prices, up to the $1000 offer cap, so as to raise market prices. Some suppliers submit "hockey-stick" offers in which a small percentage of the generator's output is priced at a high price in the event that demand is high enough to require all available generation capacity. There is one situation in which PJM will administratively raise prices day-ahead. In the condition of "maximum emergency generation," the ISO finds that day-ahead bid demand is not met by offered generation at maximum output. The ISO then takes a sequence of steps to balance supply and demand, setting the market energy price after each step to the highest offer price of any generator on-line or demand offer accepted. First, it increases the output of scheduled generation to their maximum emergency output limits. Second, it schedules generators that are designated as only available for such emergencies. Third, it drops ~~the full MW of~~ any remaining price-sensitive demand bids. Finally, it sheds load. In the final step, the market price is set at the higher of $1000/MWh or the highest offer price of a generator on-line.

In contrast, New York has established demand curves for operating reserves that administratively set energy, regulation, and reserve prices during reserve shortages. Until the reserve shortage condition, generators can, as in PJM, attempt to raise the energy price to reflect scarcity ~~through hockey stick bidding or~~ by submitting high price offers for output at what is designated as emergency levels (however, unlike PJM, almost all non-emergency supply offers are subject to automatic screening for market power and possible mitigation, as described in Section 5.8). But once the reserve shortage is reached, energy prices are administratively determined through co-optimization with the reserve markets. Currently, there are nine reserve demand curves, some for particular zones.[22] In each case, the ISO seeks a target quantity (MW) level for each type of reserve and the demand curve affects the price when the quantity falls below the target level. The curves are simply single price points or stepwise ~~curves~~ when the quantity of reserves reaches a particular level; i.e., they are not negatively sloped. The highest such price in the New York system is $500/MWh for spinning reserve. The lower quality reserves are then priced at lower levels. The prices can go no higher than addition of those demand curve prices, so that

[22] The nine demand curves are as follows. For spinning reserves the price is (1) $500/MW for eligible operating reserves at quantities less than or equal to the target level and $0/MW otherwise; (2) $25/MW for eligible operating reserves at quantities less than or equal to the target level for the Eastern region and $0/MW otherwise; or (3) $25/MW for eligible operating reserves at quantities less than or equal to the target level for the Long Island zone and $0/MW otherwise. For total 10-minute reserves (spinning and non-synchronous) (4) $150/MW for eligible operating reserves at quantities less than or equal to the target level and $0/MW otherwise; (5) $500/MW for eligible operating reserves at quantities less than or equal to the target level for the Eastern region and $0/MW otherwise; or (6) $25/MW for eligible operating reserves at quantities less than or equal to the target level for the Long Island zone and $0/MW otherwise. For total 30-minute operating reserves the price is (7) $200/MW for eligible operating reserves at quantities less than or equal to the target level minus 400 MW, $100/MW for eligible operating reserves at quantities less than or equal to the target level minus 200 MW, but greater than the target level minus 400 MW, $500/MW for eligible operating reserves at quantities less than or equal to the target level, but greater than the target level minus 200 MW, and $0/MW otherwise; (8) $25/MW for eligible operating reserves at quantities less than than or equal to the target level for the Eastern region and $0/MW otherwise; or (9) $300/MW for eligible operating reserves at quantities less than or equal to the target level for the Long Island zone and $0/MW otherwise.

even if a high availability offer is submitted in anticipation of a reserve shortage, it cannot further increase the market price of reserves.

As the comparison of PJM and New York shows, there are at least two fundamentally different approaches to scarcity pricing currently implemented. Other ISOs have proposed different methods and/or pricing parameters. Hence, there are likely to be many subsequent developments and refinements in this aspect of market design.

5.4. The Reliability Unit Commitment

While the day-ahead market with security-constrained unit commitment is now generally accepted as a design feature that improves the interface between forward markets and reliable system operations in real-time, it still presents the ISO with a few problems in this regard. These problems stem essentially from uncertainty over the physical sufficiency of the supply and demand cleared in the day-ahead market. First, as noted, when virtual supply offers are introduced into the day-ahead market, they can displace physical supply offers, potentially leaving the ISO uncertain as to the location and available capacity of generators that are preparing to operate for the next day. Second, if there is no requirement that demand bid in day-ahead, and, further, the market allows virtual demand to bid, the ISO may be uncertain as to whether the demand cleared in the day-ahead market is over- or under-scheduled with respect to the operating day. Most US ISOs have addressed these potential problems by adding an intermediate stage to the sequence of the day-ahead and real-time markets, sometimes called a reliability unit commitment.

In the reliability unit commitment, the ISO takes two primary actions. First, the ISO removes accepted virtual supply offers from the day-ahead generation offer stack to determine its prospective post-day-ahead physical supply (although the physical generators do not have to operate in real-time, they at least have a financial inducement to do so). Second, the ISO compares its own next-day load forecast with the demand cleared in the day-ahead market. If the former exceeds the latter, the ISO seeks to ensure that sufficient generation has been started up to meet load. The ISO then commits additional generators to meet its load forecast. In this fashion, the ISO enters the operating day with greater confidence that it has sufficient generation to meet demand.

Like some other design features of ISO markets, the reliability unit commitment is best thought of as a market-priced reliability procedure that could eventually be removed in the event that the demand-side of the markets become sufficiently price-responsive. In most ISO markets, it is undertaken after the day-ahead market closes but before the real-time market opens. This means that it is usually conducted in the late afternoon of the prior day, with results made available to market participants over the next few hours. As such, it can also be considered the beginning of the real-time market since, as will be discussed below, any generation committed through the reliability unit commitment will be compensated for any energy produced at real-time prices.

5.4.1. General procedures and pricing rules

As reliability unit commitment designs were introduced into the ISO markets, there was debate over the correct offer and pricing rules. The design decisions were largely focused on two questions:

- Should the reliability unit commitment act as a type of forward market, clearing a market for starting generators as well as energy, or should it simply start generation with sufficient capacity to meet forecast demand while letting the real-time market determine whether to produce energy from those units?

- If the reliability unit commitment only starts generators, but does not buy their energy, what is the appropriate pricing for such units?

The design direction taken by the ISOs on the first question was not to compute a post-day-ahead forward market for energy, but rather to minimize the cost of start-up offers by generators such that sufficient capacity is postured to meet the residual demand between the day-ahead market and the ISO's next-day load forecast. In this fashion, the ISOs could minimize the expenditures by buyers associated with the reliability commitment. If the generators are started up, but their energy is not actually used, then the ISO has not over-purchased energy that would need to be bought back at real-time prices.

Conversely, if the generators are started up and do produce energy in real-time, then they are paid for their energy through the real-time market. In fact, all ISOs reduce the costs of the reliability commitment further by applying the real-time market revenue sufficiency guarantee. That is, generators scheduled in the reliability unit commitment are not paid until the ISO has determined whether they have earned sufficient revenues in the real-time market through energy sales to cover any start-up and no-load costs incurred. The methods for allocating such uplift charges are discussed in Section 5.6.

5.4.2. Numerical example (continued)

In this section, the numerical example begun in Section 5.3 is modified to reflect the actions taken by the ISO in the reliability unit commitment. In the example, a new ISO load forecast is substituted for the demand cleared in the day-ahead market auction. These differences in these demands for each time period are shown in Table 5.8. In addition, the ISO removes the accepted virtual supply offer from the peak demand hours. The results are as follows. T8

In the off-peak period, the ISO estimates that demand will be 25 MWh higher than the demand cleared through the day-ahead market. This additional demand is met at lowest cost by the base-load generator that is already scheduled to operate during that hour. Hence, the ISO does not need to make any additional unit commitments for that period. In the intermediate period, the ISO estimates that demand will be 100 MWh higher than the demand cleared day-ahead. Again, this load is met at lowest cost by the two generators already scheduled in the day-ahead market to operate in that period.

In neither the off-peak nor the intermediate periods does the ISO need to consider the displacement of physical generation by virtual supply offers. However, in the peak period, the ISO does have to remove an accepted virtual supply offer from the day-ahead market schedule to conduct the reliability commitment. In doing so, the ISO must consider which physical generator offers are available to meet its forecast peak load, which is 10

Table 5.8. Difference between day-ahead market demand (MWh) and ISO load forecast (MWh)

	Off-peak	Intermediate	Peak
Day-Ahead Market	950	1300	1600
ISO Load Forecast	975	1400	1590
Difference	+25	+100	−10

MWh lower than the peak demand cleared in the day-ahead market, but which is being met day-ahead with the virtual supply. Both peak generators available in this example have sufficient capacity (MW) to meet the ISO's load forecast, and the ISO only needs the capacity of one of them, so the ISO's commitment decision is simply to minimize the costs of starting up one of the generators to meet the forecast demand. In this example, it is generator D, which has high energy costs but low start-up costs, that is committed by the ISO in the reliability unit commitment. By scheduling this generator, the ISO is guaranteeing that it will cover the unit's start-up costs of $100 whether or not it actually produces energy in real-time. This cost now rolls over into the real-time revenue sufficiency guarantee, as discussed in sections 5.5 and 5.6. In turn, generator D is now obligated to start up and be ready to produce energy by the peak period of the real-time market. To simplify the example, assume that generator D is a "quick start" unit that does not have a ramp rate that would require multi-period scheduling.

5.4.3. Comparison of PJM and New York ISO market rules

All the US ISOs with a day-ahead market conduct a reliability unit commitment and all procure additional capacity by minimizing start-up and no-load costs. The differences in market rules are relatively minor. PJM begins this procedure, which it calls the "secondary resource commitment," at 18:00 the prior day. This commitment uses the updated offers via the real-time market and any updated information on resource availability to meet the updated PJM load forecast. Following this initial commitment, PJM will do additional resource commitment prior to the start of the real-time market. Any changes to individual generation schedules are provided to generation owners only.

New York ISO undertakes its reliability unit commitment over multiple steps. In contrast to PJM (and the other ISOs), New York integrates its initial reliability unit commitment into its day-ahead market unit commitment. This has been described earlier. In the second pass of the five-pass unit commitment, the ISO clears additional units based on day-ahead offers to meet the ISOs forecast demand. Like the other ISOs, the objective is to minimize start-up and no-load costs. However, because there is no guarantee that sufficient supply will be offered into the energy market day-ahead, following the day-ahead commitment, New York ISO conducts a resource evaluation called "forecast required energy for dispatch." In this evaluation, it solicits any additional offers that were not submitted to the day-ahead market or represented in day-ahead schedules. At the start of the real-time market, the ISO continues to do supplemental evaluations to ensure that sufficient offers are available for dispatch. Similarly to other reliability unit commitments, units committed are eligible for start-up and no-load payments and for the revenue sufficiency guarantee.

5.5. Real-Time Market

The real-time, or dispatch, market encompasses both an auction market for energy and ancillary services and any submitted MW deviations from day-ahead self-schedules and bilateral schedules. The spot auction is a purely "physical" market in that all sales and price-sensitive demand bids cleared through the market embody requirements to produce or curtail consumption. In the real-time auction market, demand that did not clear in the day-ahead market purchases energy and ancillary services from generators that have been started either through the day-ahead market (and have surplus capacity), the reliability unit commitment or through ongoing real-time commitments. If excess demand cleared through the day-ahead market, then the real-time market calculates the prices at which that demand must be "sold back." Similarly, if supply clears through the day-ahead market, whether backed by a physical generator or a virtual offer, but does not perform in real-time, its output must be "bought back" at the real-time price. In markets with locational marginal pricing, the real-time market also determines marginal congestion and marginal loss charges for any transactions in real-time. These charges are calculated based on the total dispatch, but settled against deviations from the day-ahead market, as shown in Section 5.8.

5.5.1. *Market procedures*

Offers and bids into the real-time market include those submitted into the day-ahead market but not accepted in that market and new ones submitted subsequent to the day-ahead market. ISOs differ over when offers and bids are due, as summarized in Table 5.3. The major difference is between markets that collect all real-time offers and bids the prior day (PJM) and those that allow them to be submitted up to some time just prior to the hourly market in question (all other US ISOs). The price components and physical parameters required for offers into the real-time market are generally the same as the offers into the day-ahead market, as summarized in Table 5.7.

The real-time market begins market clearing operations at 00:00 of the operating day and closes at 24:00 of the same day. As with the day-ahead market, by convention it is operated as an hourly market in the US ISOs. Hence, there is a market for each hour (e.g., 12:00–13:00) and typically a single hourly integrated price at each location, even if the dispatch price is calculated on a 5–15 minute' basis.

5.5.2. *Market pricing*

There are similarities and differences between the day-ahead market and the real-time market with respect to the calculation of LMPs for energy. The similarity is that once calculated, LMPs clear the energy market. However, the procedure for calculating the prices is quite different. Unlike the day-ahead auction market, the real-time market typically has two parallel programs running to schedule and dispatch generators. The first begins with the day-ahead and reliability commitments and conducts unit commitment and de-commitment over the operating day, with a look-ahead usually of a few hours. This unit commitment typically only has to consider peaking units that were not scheduled by the day-ahead commitment or whose schedules need to be adjusted. The second is the dispatch program, which takes the commitment decisions as given and adjusts generator output on a 5-minute basis to achieve an optimal dispatch and set LMPs. The better the integration between these two scheduling programs, the more efficient is the dispatch.

Also unlike the day-ahead market, in real-time, the ISO both sends its computed dispatch prices and outputs to generators every few minutes (often 5 or 10 minutes) for the subsequent time period and then meters the actual output a few minutes after that. There are thus two possible ways to set real-time LMPs: on the basis of the *ex ante* dispatch results or by using the metered outputs to calculate prices *ex post*. *Ex ante* real-time pricing has the advantage that it is reflective of the ISO's optimal dispatch for the next time period. However, it generally requires penalties for "uninstructed" deviations (i.e., deviations that take place contrary to system operator instructions), because if a generator knows the anticipated price for the next 5–10 minutes, it may choose to over or under-produce in violation of the ISO's optimal dispatch. Such deviations may cause costs to other parties, by requiring other generators to ramp up or back down (as regulation), or affect reliability. *Ex post* pricing relies on incentives for generators to follow their dispatch instruction; generally, if a generator over-produces, it lowers the locational price. But because of transmission network effects, some ISOs do not find that *ex post* pricing offers them sufficient control over the system and rely on *ex ante* prices along with penalties to maintain an optimal economic dispatch.

Another difference with the day-ahead market is that in real-time, while dispatch instructions and LMPs are calculated on a 5–10 minute basis, for convenience financial settlement takes place against an average hourly price. Hence, some hourly real-time LMPs do not cover the offer price of some units dispatched during the hour. Any difference, as in the day-ahead market, is made up through the revenue sufficiency guarantee.

5.5.3. Markets for regulation and reserves

Real-time markets for regulation and reserves cannot be entirely operated in the context of the 5-minute dispatch because they require commitment decisions and re-scheduling based on the dispatch points that generators have moved to over the operating day. Hence, these are operated as hourly markets and typically re-adjusted in each hour based on new offers and the results of the energy dispatch prior hour. ~~Hence,~~ the auction market for these products runs parallel to the energy dispatch auction. However, for any energy produced by a generator providing regulation or reserves, the price paid is the real-time energy price.

5.5.4. Scarcity pricing

Scarcity pricing in the real-time market typically follows similar principles to that in the day-ahead market. Again, because spot demand is largely non-price responsive, and more so in real-time than day-ahead, most ISOs again use a reserve shortage as a proxy for market shortage in real-time (even if no demand curtailments take place). Two significant differences between the two markets are in the trigger for scarcity pricing and financial settlement. In real time, system operators will undertake many non-market measures to avoid running short of reserves. Hence, there can be some ambiguity about when scarcity pricing is triggered. Another difference is that due to the averaging of real-time hourly prices, high scarcity prices for a few minutes during the hour will be averaged over the hour, resulting in a diluted price signal to the market.

5.5.5. Congestion management

Unlike the day-ahead market, real-time congestion management requires physical re-dispatch throughout the day. The system operators use LMPs and dispatch instructions to resolve congestion, but can also take numerous non-market actions. For example, in PJM, prior to generation re-dispatch, the operator takes all available "non-cost" measures to resolve congestion, including PAR adjustments, transformer tap adjustments, MVAR adjustments, switching capacitors/reactors in/out-of-service, switching transmission facilities in/out-of-service, and curtailing transactions that have indicated that they are "not-willing-to-pay" congestion. After these non-cost measures are completed, the system operator sets a "threshold" for each individual congested transmission element, usually 95–100% of the facility rating, which must be respected by the unit dispatch software. This threshold is then re-adjusted by the PJM dispatcher as conditions change.

5.5.6. Numerical example (continued)

As noted, the real-time market is a residual market, in which deviations from day-ahead market schedules are settled at real-time prices. ~~In practice, the ISO will recalculate the entire market solution, setting day-ahead market offers at their scheduled quantities and substituting new supply offers upon request for any offers not accepted through the day-ahead market and adding any that were submitted after the close of the day-ahead market.~~ To maintain the simplicity of the numerical example, only the supply offers submitted into the day-ahead market but not accepted in that market or the generator committed in the reliability unit commitment are considered for additional real-time demand. The prices and quantities in their supply offers are assumed not to have changed. What does change in the example is the real-time, or physical, demand. The results of each demand scenario are shown in the Figs 5.3a–c. As shown in Table 5.9, actual demand is 5 MW T9 lower in the off-peak period than the quantity cleared day-ahead, but is 150 MW and 20 MW higher than the latter for the intermediate and peak periods, respectively. As a result, the physical dispatch and nodal prices are different from the day-ahead market.

In the off-peak period shown in Fig. 5.3a, the base-load generator is dispatched at a level F3a 5.76 MWh lower than the quantity settled in the day-ahead market. Since the generator remains marginal, the price at its bus remains the same as the day-ahead market off-peak period. Hence, the generator has to "buy back" 5.76 MW × \$15/MWh = \$86.45. Buying back this forward position does not cause the generator to lose money, since it did not produce any physical power day-ahead. Similarly, the demand at bus 3 bought excess power day-ahead, which it now "sells back" at a slightly lower price than the day-ahead price at its location, 5 MW × \$17.29/MWh = \$86.45. The differences in the nodal price at bus 3 between day-ahead and real-time are due to differences in losses. In this case, the ISO remains revenue neutral.

In the intermediate period shown in Fig. 5.3b, generator C, which was not dispatched in the day-ahead market due to the virtual supplier, is committed because demand is higher than anticipated by the day-ahead market. Remember that the reliability unit commitment postured generator D to potentially provide energy in the peak period, and in doing so committed to paying its start-up. However, by the intermediate period of the real-time market, using its look-ahead unit commitment software, the ISO is aware that both generators C and D will be needed for the peak period. Hence, it is economic to use generator C rather than generator D for the intermediate period. In fact, had both generators not been needed for the peak period, it would have been economic to

Table 5.9. Results of real-time market simulations and deviations from the day-ahead market

Scenario	Demand (MWh)	Supply (MWh)					Price at bus (\$/MWh)		
		A	B	C	D	E	1	2	3
Off-peak	945	1013					15.00	16.13	17.29
Intermediate	1450	1288	250	47			15.00	52.80	40.00
Peak	1620	1288	250	200	17	0	15.00	69.75	50.00
Deviations from the day-ahead market:									
Off-peak	−5	−6			0		0	−0.02	
Intermediate	+150	+61	+60	+47			0	+32.80	+19.64
Peak	+20	0	0	+200	+17	−197	0	+15.26	+9.00

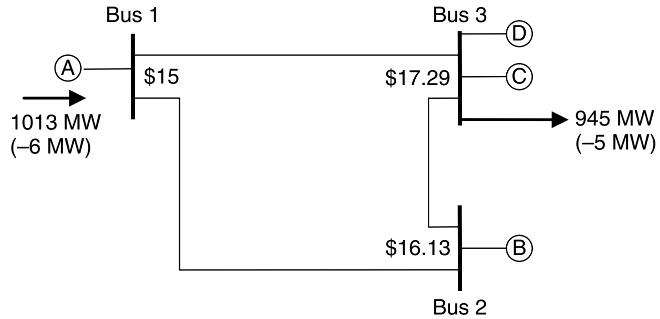

Fig. 5.3a. Off-peak scenario.

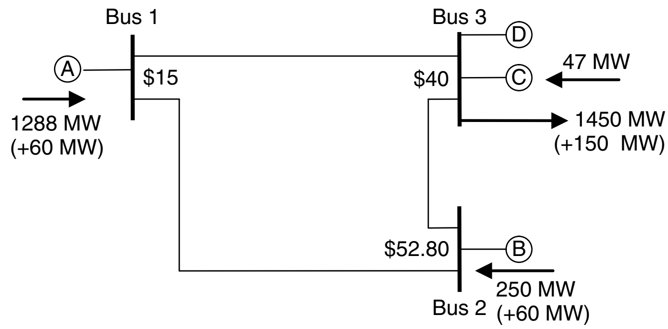

Fig. 5.3b. Intermediate scenario.

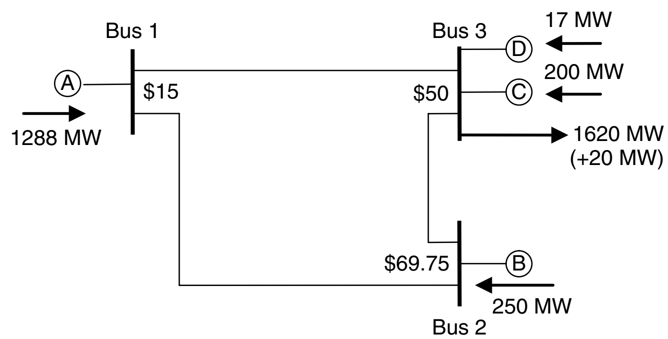

Fig. 5.3c. Peak scenario.

"decommit" generator D following the reliability unit commitment and only use generator C due to its lower energy price. That is, the savings due to lower energy prices would have more than offset the higher cost of starting up generator C.

In the peak period shown in Fig. 5.3c, both peaking generators are dispatched, because peak demand is higher than was cleared in the day-ahead market and also higher than the

ISO forecast in the reliability unit commitment. As noted, the ISO committed generator D in the reliability commitment, and in doing so was obligated to pay the generator its start-up costs regardless of whether it provided energy. Now that the generator has been dispatched to provide energy, the ISO must calculate whether its revenues are sufficient to cover its start-up costs. These calculations are shown in Section 5.6.1. In addition, the virtual supplier must "buy back" its position in the period in which it sold virtual energy day-ahead but at the real-time price. In this example, the virtual supplier ~~sells~~ back 197 MW at bus 3 at the real-time price of \$50.00/MWh. Hence, the virtual supplier owes back \$9850, for a net loss between day-ahead and real-time of (-\$50 + \$40) × 197 MW = -\$1970. If the virtual supplier's forecast that the real-time price would be \$41.00/MWh had been correct, then it would have owed back −\$7879, for a net profit between day-ahead and real-time of (\$41 - \$40) × 197 MW = \$197.

5.5.7. Comparison of PJM and New York ISO market rules

As with the day-ahead markets, the PJM and New York real-time markets both calculate LMPs for energy and zonal prices for several ancillary services. However, there are again some interesting differences in market design and procedures.

5.5.7.1. General dispatch procedures and energy market rules

Beginning with the energy markets, the two ISOs have different offer price rules and trading deadlines. In PJM, the real-time market offer period begins immediately following the close of the day-ahead market at 16:00 and closes at 18:00 the prior day. In addition, like the day-ahead market, only a single supply function offer can be submitted for all 24 hours of the real-time market. In New York, offers and bids are only due 75 minutes before the operating day hour, and the offer price can vary for each hour.

Unlike the day-ahead auction, the real-time auction schedule is the result of the interaction of market functions with several constantly updated forecasting and system monitoring systems. These are worth examining in some detail. In PJM, with all offers submitted, the system operators begin calculating *ex ante* dispatch instructions for the operating hour using its unit dispatch system. Data from three software systems feeds into the unit dispatch analysis: the energy management system calculates the load forecast, area control error, steam deviation, constraint data, unit sensitivities, and state estimator output; another software system updates unit outage data; and market systems software provides generator offers and data on dispatchable transactions (PJM, 2007b, d). In addition, as described below, a parallel program clears markets for regulation and spinning reserves using the unit dispatch data.

The PJM energy dispatch auction subsequently clears using the most current data from the various software subsystems. The unit dispatch system calculates a dispatch solution including dispatch rates and generator unit MW every 5 minutes for a look-ahead period. The dispatcher must approve the solution before the data is sent to utilities and generators. The ISO then solves for a state estimator solution every 5 minutes to estimate the actual MW injections and withdrawals at buses. Actual schedules for external transactions are then included and the ISO sets real-time prices on an *ex post* basis. There are no penalties for deviations.

In New York, as with the day-ahead market, in the real-time market, energy, regulation, and operating reserves are co-optimized. The calculation of real-time schedules and prices is implemented through two primary commitment and dispatch programs that exchange data with each other: the real-time unit commitment program, which conducts a 2.5-hour look-ahead with commitment decisions made on a 15-minute basis, and the real-time dispatch program, which establishes a 5-minute dispatch and calculates market prices on a 5-minute basis. The real-time unit commitment begins with the day-ahead schedule and either commits or decommits units from that schedule based on offers submitted following the day-ahead market, changes in resource availability and updated load forecasts. Its commitment schedule is binding for units that need to be started up or moved to a dispatch point over a 30-minute look-ahead and advisory for units committed beyond that horizon. The commitment decisions are then passed to the real-time dispatch program, which can adjust the output of committed generators to determine an optimal dispatch and calculate prices.[23]

The dispatch produces *ex ante* binding prices and quantities every 5 minutes for the next 5- minute period and provides four additional advisory prices and quantities spaced over 5–15 minutes up to 50–60 minutes ahead, depending on the initial period.

5.5.7.2. *Markets for regulation and reserves*

To establish the real-time dispatch, the ISOs have different methods for co-optimizing the offers for regulation and operating reserves with those for energy. As noted, the New York ISO operates a two-settlement system for regulation and operating reserve (NYISO, 2006). The basic rules for these markets were described above in the section on the day-ahead market. New York buyers purchase most of their regulation and reserve requirements through the day-ahead market. The ISO commits additional resources to provide these services in real-time if insufficient MW were cleared day-ahead, units that were scheduled day-ahead are not available in real-time, or additional MW are needed over the day-ahead forecast (only for regulation, which is a function of actual load). The offer rules for the New York real-time markets for operating reserves are largely the same as those for the day-ahead market, with the exception that all availability offers into the real-time market are assigned a price of $0/MW, meaning that their payments will be based on opportunity costs relative to their energy offer.

In contrast to New York, PJM's ancillary service markets clear after the day-ahead market, and before the real-time market (PJM, 2007b). PJM operates hourly real-time markets for two ancillary services, regulation and spinning reserve (called synchronous reserve).[24] Suppliers include eligible generators and demand resources. Demand resources are currently limited to providing 25% of the regulation requirement. There are two types of spinning reserve resources: Tier 1 and Tier 2. Tier 1 is any incremental spinning reserve that is already available through the energy dispatch; i.e., from a generator that is already operating and has additional capacity available through ramping. Tier 2 is spinning reserve from

[23] The dispatch program's procedure for normal (i.e., non-emergency) time periods is a three-pass approach. The first pass determines an initial set of binding physical schedules for generators that result from co-optimized minimization of the cost of energy, regulation, and operating reserves. This pass assumes that all fixed block units that have been committed by the real-time commitment are at their upper operating limits, whereas all other dispatchable capacity that has not been committed is flexible. The second pass then relaxes the constraint on loading fixed block units and allows them to be flexibly loaded. This pass determines whether the least-cost solution can be found through inflexible or flexible loading of such units. The third pass then calculates LMPs with the optimal mix of inflexible or flexible block loaded units. However, the third pass does not change the physical schedule determined in the first pass.

[24] The regulation market began on June 1, 2000; the spinning reserve market began on December 1, 2002.

Els UK Job: CDI ch05-i047172 25-10-2007 7:33p.m. Page:221 Trim:165×240MM Float:Top/Bot TS: Integra, India

generators that are synchronized to the grid but which need to be dispatched to a different operating point than they would be through the energy dispatch (including generators started-up to produce reserves). Regulation is procured in two separate zones in PJM, while spinning reserve is procured in four zones, with zonal prices if transmission constraints separate the zones and a single price when they do not.[25] Each load-serving entity must buy a pro-rata share of these services. As with energy, buyers can self-schedule using their own generators, contract with a third party, or purchase through the PJM market.

Supply offers for regulation and Tier 2 reserve suppliers are due to PJM by 18:00 the prior day and the price cannot vary by hour of the day.[26] Also, any units listed as available for these services but without an offer price are entered into the markets as price takers (i.e., with their prices offers set to zero). However, physical parameters can be changed up to 1 or 2 hours before the dispatch hour, as described below.

In PJM, the process for scheduling ancillary services and calculating the dispatch similarly begins 1–2 hours prior to the dispatch hour. The first step in real-time market pricing is to calculate the prices for regulation and Tier 2 reserves, which are actually done prior to the real-time energy market and are thus *ex ante* rather than *ex post*. Since most units will provide both energy and ancillary services, PJM's market deadlines for submitting physical parameters for both regulation and spinning reserve are within 1–2 hours of the operating hour. The data on ancillary services (including price offers) is then evaluated by PJM using its unit dispatch system software. For regulation, the final regulation capability (MW) above and below the regulation midpoint and the regulation maximum and minimum values (MW) must be finalized 1 hour prior to the operating hour. For spinning reserves, ramp rates and maximum reserve MW are due 2 hours prior to the dispatch hour. This information is then used to estimate the Tier 1 reserve schedules, which are posted 90 minutes prior to the dispatch hour. Reserve availability and offer quantities for Tier 2 resources are due by 1 hour prior to the dispatch hour. Self-schedules are also due by 1 hour prior, with exceptions for units substituted for others that have become unavailable and for units that have only become available during the dispatch hour.

The pricing of ancillary services is then conducted through a co-optimization of forecast energy prices for the hour with the offers and parameters submitted for regulation and spinning reserves. The forecast LMPs are the result of a 1-hour look-ahead provided by PJM's unit dispatch tool. For regulation, PJM calculates a supply stack that reflects each regulation units offer and any opportunity costs incurred by not producing energy. The highest merit order unit price becomes the regulation market clearing price for the hour.

Similarly to regulation, for the Tier 2 reserves, PJM's objective is to calculate a supply stack that reflects a Tier 2 unit's offer price for standing by on reserve as well as any opportunity costs that it might incur by not providing energy (a demand resource has an opportunity cost of zero). The formula is as follows:

> Resource merit order price ($/MWh) = Resource synchronized reserve offer + estimated resource opportunity cost per MWh of capability + energy use per MWh of capability.[27]

---

[25] The PJM regulation requirement is 1% of the PJM peak load for the day.

[26] For regulation, this offer restriction was justified on the basis of market power concerns; see discussion in Section 5.8.

[27] PJM applies different formulas for the estimated resource opportunity cost. For condensing combustion turbines, this opportunity cost is calculated as the absolute value of the difference between the unit's energy offer price and the forecast LMP at the unit's bus multiplied by the unit's MW capability

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T. W:30PC open recto 1 Color 49 Lines

The price of Tier 2 spinning reserves in each zone is the highest resource merit price for the operating hour. The prices for regulation and spinning reserves are posted no later than 30 minutes prior to the operating hour.

Tier 1 reserves are priced not for capacity on reserve but on the basis of actual output in response to a reserve call. The price is a $50 premium over the LMP for energy. In the event that spinning reserve units are called to provide energy and do not perform, Tier 1 units are credited for the MW that they provide but are not penalized otherwise, while Tier 2 units have to repay for any non-performance by providing the MW shortfall for the next 3 consecutive, same peak days.

5.6. The Revenue Sufficiency Guarantee

As described above, participants in the ISO markets are subject both to market incentives and to operational instructions that may be required to maintain system reliability. To establish the appropriate incentives for market sellers and buyers to follow both auction schedules and operational instructions, there are two types of rules. First, there is a "revenue sufficiency guarantee" that any multi-part offer accepted by the auction will be fully compensated through additional payments if market revenues are not sufficient to meet its offer requirements. These additional payments are then billed to buyers as an "uplift" – an *ex post* charge assigned on some averaged basis. This rule will be the subject of this section. There is a second type of rule, related to the first but less frequently needed, which is that any participant that has sold or bought through the markets and is then requested by the ISO to change its schedule or physical position for reliability reasons will not lose money in doing so.[28]

In the early phases of market design, the question of how to calculate and allocate revenue sufficiency guarantee charges was a quite heavily contested issue.[29] The simplest proposition was that such uplift should be paid on a load-weighted share basis by all demand. However, for merchant suppliers that were interested in primarily transacting through the forward contract markets, there was a concern that allocating the uplift to buyers through the spot markets would provide a financial advantage to sellers into the spot markets. That is, a generator seeking to create a short-term bilateral contract would have to incorporate its start-up costs into the forward contract price, whereas spot sellers could have their start-up costs spread out over the buyers in the ISO market. Further, bilateral buyers might thus double-pay if they both paid for start-up via a forward contract and for other spot sellers as part of an uplift charge. Some sellers thus asked that their contracted purchasers be insulated from the market-wide uplift charges. The position taken by the ISOs and supported by FERC (e.g., FERC, 1999) was that generators would be started up to provide both energy and ancillary services, such as regulation and operating reserves;

---

with this number then divided by the unit's synchronized reserve capability. For non-condensing units, this opportunity cost is calculated as the absolute value of the difference between the forecast LMP and the price estimated for the unit's set point to provide its assigned quantity of synchronized reserve multiplied by the quantity of synchronized reserve provided. Finally, the energy use component of the price is calculated as the forecast LMP multiplied by the MW of energy use divided by the synchronized reserve capability.

[28] For example, a generator may be asked to ramp down even though the prior real-time LMP at its bus indicated that it should increase production. These types of problem are often related to difficulties in integrating the market price calculation with the short-term operational decisions.

[29] See, e.g., discussion of New York ISO's rules in FERC (1999), pp. 42–3.

since the latter were reliability services that buyers could not easily dis-aggregate from energy, there was a rationale for allocating the uplift to all buyers on a total demand basis.

Another design issue concerns the assignment of such uplift to virtual transactions. As noted above, a virtual supplier "sells" energy in the day-ahead market and then "buys it back" in the real-time market. The virtual seller may displace some physical generation offers in the day-ahead market that are then re-scheduled through the reliability unit commitment. In theory, the start-up and stand-by of the physical units may embody some costs that are then passed to the revenue sufficiency guarantee uplift, either day-ahead or in real-time.[30] In some cases, such virtual positions may cause uplift to be shifted from one party to another.[31]

Whether such *ex post* uplift charges should be allocated to the virtual suppliers is a contested design issues in the US ISO markets (see, e.g., Hogan, 2006). There is no simple and accurate way to calculate the actual price impact of virtual sellers on real-time market uplift charges. The most accurate (in a static sense) would be to re-solve the day-ahead and real-time market sequence with and without virtual transactions represented as a means to quantify the difference. This might be a relatively simple analysis for the day-ahead market, requiring one pass of the unit commitment algorithm, but obviously in real-time, it would require re-solving the dispatch algorithm for every 5 -minute market clearing to determine which generators were dispatched to a different level due to the removal of the virtual suppliers.

The simplest method is to treat all virtual supply MW as effectively demand in the real-time market and to assign uplift charges to such virtual suppliers proportionally to their share of total MW. This approach does not ~~treat~~ the impact of virtual suppliers on the ~~margin~~. For example, some virtual supply might be accepted day-ahead but not result in any physical generation being scheduled that would not have been otherwise, but would nevertheless face uplift charges on per-MW basis. In the face of such alternatives, another option is to not charge virtuals any uplift on the basis that the economic benefits that result from market price convergence outweigh any cost shifts that they might cause (e.g., as recommended by Hogan, 2006). FERC has generally supported the principle of "cost-causation" – that entities that create costs should be billed for them – and the simple approach of assigning virtual supply uplift charges on a per-MW basis, although it has not precluded more the complicated analysis of virtual supply's ~~marginal price~~ impact.

5.6.1. Numerical example (continued)

This section applies the principles of revenue sufficiency to the numerical example. In Tables 5.10 and 5.11, the total revenues that each generator earns in the day-ahead and T10,T11
real-time markets are compared to its offer price requirements for energy and start-up

[30] Similarly a virtual demand bid that clears the day-ahead market may cause a physical generator to be scheduled that would not have been otherwise. In the reliability unit commitment, the ISO uses its own load forecast so that virtual demand can be removed from the schedule.

[31] For example, if there is no charge to virtual suppliers, a vertically integrated utility that is an actual consumer of power, i.e., will have metered energy withdrawal in real time, can also shift uplift charges from day-ahead to real-time, if that is financially advantageous, by taking a virtual seller position. To do so, it substitutes virtual supply for its actual generation day-ahead, thus reducing any revenue sufficiency uplift paid by its day-ahead load. Its physical generators are then scheduled through the reliability unit commitment or in real time, creating revenue sufficiency uplift. However, since its day-ahead load has not deviated, all the uplift is billed to net real-time load.

Table 5.10. Day-ahead market calculation of revenue sufficiency

	Revenues ($)				Offer Requirements ($)					Revenue Sufficiency
	Off-peak	Intermediate	Peak	Total	Off-peak	Intermediate	Peak	Start-up	Total	
Generator A	15 275	18 411	19 316	53 002	15 275	18 411	19 316	–	53 002	0
Generator B		3 793	13 623	17 416		3 793	5 000	1 000	9 793	7 623
Virtual E			8 076	8 076			8 076	0	8 076	0

Table 5.11. Real-time market calculation of revenue sufficiency (including reliability unit commitment)

	Revenues ($)				Offer Requirements ($) (including reliability unit commitment)					Revenue Sufficiency
	Off-peak	Intermediate	Peak	Total	Off-peak	Intermediate	Peak	Start-up	Total	
Generator A	−87	905	–	819	−87	905	–	–	819	0
Generator B		3186	–	3186		1207	–	–	–	1979
Generator C		1879	10 000	11 879		1879	8000	2000	11 879	0
Generator D			849	849			849	100	949	−100

shown in Table 5.4.[32] The total revenues for each generator in each energy market are the generator's output (MWh) in each period multiplied by the LMP ($/MWh) at its location, and summed over all periods. This revenue is shown in column 5 of Tables 5.10 and 5.11 (the source data is shown in Figs 5.2a–c and 5.3a–c, and Tables 5.5 and 5.9). The total offer requirement for each generator is equal to the generator's start-up offer price and the sum of the generator's output in each period multiplied by its offer price ($/MWh). This requirement is shown in column 10 of Tables 5.10 and 5.11.

All generator offers and the virtual offer are included in Tables 5.10 and 5.11, although not each of these offers is eligible for the revenue sufficiency guarantee in the example. For example, the base-load generator A is not assumed to have submitted start-up costs to the ISO because it is not starting-up in the period of the daily auction. Similarly, the virtual offer E is not eligible because it has no start-up offer (and is not allowed to submit one).

The final column in Tables 5.10 and 5.11 shows whether each offer is revenue sufficient. In the day-ahead market, shown in Table 5.10, each supply offer is revenue sufficient, and generator B makes revenues that exceed its offer requirements. However, in the real-time market, shown in Table 5.11, generator D does not earn sufficient revenues in the energy market to cover its offer prices for energy and start-up. It is thus owed $100 by the ISO. As noted, because the ISO is revenue neutral, this uplift payment is charged to some set of market participants. For example, it could be charged to real-time demand. Table 5.9 shows that in the intermediate and peak periods, 170 MWh of demand was present in real-time that was not cleared day-ahead (ignoring the −5 MWh deviation in the off-peak period). Hence, spreading the $100 over this demand on an averaged basis would result in an approximately $0.59/MWh additional charge to each 1 MWh purchase of energy in real-time. If the −197 MWh deviation caused by the virtual supplier was added to the real-time demand, as it is in some ISOs, then the per MWh uplift would be approximately $0.27/MWh.

5.6.2. Comparison of PJM and New York ISO market rules

The ISOs allocate revenue sufficiency guarantee uplift in roughly the same fashion, but with some differences. PJM collects the uplift for the revenue sufficiency guarantee for energy as a part of its charges for providing operating reserves, presumably because start-up and no-load payments are required for units that are turned on to provide reserves AU5 (PJM, ~~2006a~~; see also discussion in PJM, 2004). PJM calls this payment "operating reserves charges" and applies them in both the day-ahead and real-time markets. PJM allocates such charges day-ahead to the day-ahead demand, including accepted virtual bids (called decrement bids in PJM), and exports; in real-time, any additional charges are allocated to deviations from day-ahead schedules, including virtual supply (called increment bids in PJM) and demand. Similarly to PJM, New York ISO collects this uplift on a pro-rata basis from all wholesale buyers, but through a per unit charge for transmission scheduling. New York ISO also charges virtual supply for any incremental costs that such transactions cause through the real-time revenue sufficiency guarantee (NYISO, 2005).

5.7. Pricing and Settlement of Marginal Congestion and Losses

Heretofore, locational marginal pricing of energy has been discussed at a general level, noting that such prices in the ISOs typically reflect the effect of both ~~the~~ marginal congestion

[32] In the ISO markets, the revenue sufficiency guarantee extends also to the "no-load" component of the energy offer.

and ~~the~~ marginal losses. In this section, more detail is provided on how the ISOs calculate these components of LMPs and how they collect and dispose of the marginal congestion and marginal loss surplus payments. The computational issue is worth discussing because it relies on certain technical assumptions and the results are sometimes misunderstood. When the transmission constraints that are examined here – capacity limits on transmission facilities and loss factors – affect LMPs, the ISO almost always collects surplus payments through the energy auction. Because the ISO is a revenue neutral organization, it must dispose of that surplus, i.e., refund it to market participants in some fashion. The rules that are devised for such refunds are somewhat different in each ISO. AU6

5.7.1. Computation of marginal congestion and marginal losses

As mentioned in Section 5.3.2, an LMP can be disaggregated into three components, using methods discussed below (Schweppe et al., 1988):

1. The price of energy at a slack or reference bus located on the ISO network;
2. The marginal congestion cost associated with delivering energy from the reference bus to another node;
3. The marginal loss cost associated with delivering energy from the reference bus to another node.

The LMP is the sum of these three factors. ~~These LMP components have the following properties for purposes of pricing of transmission usage and refunding of auction surpluses:~~

1. ~~The sum over all nodes of all marginal congestion costs at each node multiplied by the (net) MW withdrawn or injected at that node is the total marginal congestion surplus.~~
2. ~~The sum over all nodes of all marginal loss costs at each node multiplied by the (net) MW withdrawn or injected at that node is the total marginal loss surplus.~~
3. The difference in the congestion or loss component between two nodes on the system, A and B, equals the marginal congestion or loss charge associated with injecting at A and withdrawing at B (i.e., a bilateral schedule).
4. ~~The difference between the sum of the marginal congestion surplus and marginal loss surplus and the total surplus earned by the ISO is equal to the negative of the total system losses times the energy price at the reference bus.~~

There are several ways to calculate the LMP components. The following method is the most common. First, the energy component is defined as the LMP at some (arbitrary) "slack" or "reference" bus, or as a weighted sum of LMPs over a set of "distributed" slack buses assuming a fixed set of proportions. The loss component of a bus LMP is the cost of the marginal losses resulting from increasing the load at that bus by 1 MW, assuming that the entire load increase, including incremental losses, are met from the slack bus (or distributed slacks according to the assumed proportions). The loss component may be positive or negative, depending on whether losses or increase or decrease as a result of the load increase. Finally, the congestion component is the difference between the bus's LMP and the sum of the energy and loss components. Note that these components depend on the arbitrary choice of a slack bus; what is not arbitrary is their sum – the LMP – at each location.

In theory, congestion-only payments to financial transmission rights depend on this arbitrary choice of slack. However, ISOs have generally gotten agreement from stakeholders on the definition of the slack buses; for instance, one approach is to use a distributed slack based on the average load distribution (e.g., California ISO, 2005). This is equivalent to assuming that an addition of 1 MW of load at any bus is met by decreasing the original loads at all buses by the same percentage. An advantage of this definition is that the load-weighted value of the loss component is then, in theory, equal to zero.

The definition of the LMP components allow for a decomposition of the total surplus earned by the ISO into its loss and congestion components. Multiplying the net withdrawal (load minus generation) at each bus by the congestion component and then summing over all buses yields the congestion surplus; then subtracting this from the total surplus gives an estimate of the loss surplus. Note that the loss surplus will then have two components: the sum of the net withdrawals times the loss component minus the total system losses times the energy component.

An alternative approach for separating the two surpluses is to define the congestion surplus as the sum of the flowgate prices [dual variables for transmission component capacity constraints in the dispatch solution; see Eqs(2) and (3) of Appendix 5A] times the flow, and then the loss surplus is the congestion surplus subtracted from the total surplus. However, the total loss surplus defined in this manner cannot be disaggregated into a bus-by-bus loss component, and so this definition has not been adopted by any ISO.

5.7.2. Disposal of marginal congestion charge surplus

By definition, the ISO collects congestion surplus payments from buyers whenever transmission capacity constraints bind in the auction. This is because transmission congestion prevents the cheapest generators from operating at full capacity prior to the dispatch of more expensive units. The ~~auction~~ surplus occurs when the ISO collects more from buyers than it owes to sellers due to such congestion. Although the total dollar amount that the ISO collects in surplus congestion charges varies over time and among the ISOs, it is
AU7 generally measured as being under 10% of total market participant expenditures.[33]

The primary mechanism for refunding these congestion surpluses to ISO market participants is through the assignment of financial transmission (property) rights, subject to the requirement of simultaneous feasibility (which ensures that the ISO will be revenue adequate).[34] In general, ISOs either allocate transmission rights directly to certain market participants (in the United States, the allocation is to "load-serving entities," which is typically defined as the party that has the contract to serve retail demand) or conduct auctions for the rights and then provide the auction revenues to the eligible market participants (again, in the United States, the load-serving entities). However they are obtained, these transmission rights collect most of the congestion surpluses that the ISO collects on an hourly basis in the day-ahead market (see, e.g., PJM, 2007a, p. 266). Any surplus congestion rents remaining after transmission rights are settled represents uses of the grid by parties that do not hold transmission rights but pay congestion charges. If there is any congestion surplus left over at the end of the year (or the period defined in the market rules), each ISO has rules defining how it is disposed. Some of these rules are discussed below.

[33] For example, PJM reports that total congestion costs have ranged between 7 and 10% of total billings between 2002 and 2006 (PJM, 2007a). See also discussion on congestion metrics in Chapter 4.

[34] That is, the ISO will collect sufficient congestion charges in each market in which financial transmission rights are settled to pay all the outstanding rights. This rule holds as long as the system topology used in the simultaneous feasibility of the rights remains the same in actual market (see ~~Hogan~~ et al., 1997).

5.7.3. Disposal of marginal loss charge surplus

The calculation of marginal losses provides the market with a more efficient dispatch than would be the case if locational prices only reflected congestion. However, because losses are usually represented as a quadratic function of line loadings (e.g., Schweppe et al., 1988), the marginal loss charge between two locations will be greater than the average loss charge. The quadratic function implies that the average cost of losses is roughly one-half the marginal cost. Hence, the ISO will always collect surplus marginal loss payments.

The disposal of the marginal loss surplus has been a controversial issue in market design, more for reasons of equity than efficiency. In general, any method for disposal of this surplus will support efficient scheduling by a particular market participant as long as the method leaves the participant indifferent between accepting the ISO schedule or dispatch and undertaking an alternative, inefficient schedule or dispatch to obtain loss refunds (e.g., by changing its supply offers or self-scheduling).[35] However, equity would suggest that the refund method is not entirely arbitrary, since it could involve unfair transfers among market participants based on pre-existing historical contracts.[36]

Market participants concerned about the uncertainty of the relationship between marginal loss charges and their share of their refund often have sought a tradable loss hedging right similar in principle to financial transmission rights for congestion. Because of the non-linearity of line losses, designing such a tradable loss right has not been straightforward. Losses are an example of diseconomies of scale or super-additive costs, such that $K(y_1) + K(y_2) \geq K(y_1 + y_2)$. Therefore, there is no simple way to decentralize trading of losses, since loss is a function of power flows (unlike transmission capacity, which is assumed to be independent of power flows in the transmission rights model discussed earlier). However, there is less uncertainty about losses than about congestion charges. Therefore, even though average losses through the year can be approximately as costly as transmission, the risks to transactions are less. Therefore, financial transmission rights that cover just congestion costs are likely to cover most of the risks that market participants care about.

5.7.4. Numerical example (continued)

In the examples given in Sections 5.3.6 and 5.5.5, the total LMP surplus collected by the ISO was calculated as the difference between payments by buyers and payments to sellers. In this section, that total surplus is disaggregated into congestion surpluses and loss surpluses for two of the day-ahead examples: the off-peak and intermediate scenarios

[35] For example, consider a buyer that has the choice between scheduling generator A or B to serve load at C. A has a lower energy offer price (and actual marginal cost) than B but a higher loss charge than B associated with deliveries to C, resulting in a higher LMP at the load bus than if B was scheduled. The efficient schedule is to dispatch B. As long as the loss refund method makes the buyer indifferent between accepting the scheduling of A or B, it will result in the efficient scheduling of B. In this simple case, if the loss refund exactly matched the loss charge for scheduling A, then the buyer would have the incentive to self-schedule A inefficiently.

[36] In some US regions, notably the Midwest ISO, some generator siting and contractual decisions made prior to the start of the ISO market were not fully reflective of actual losses (implying some cost shifting). In such regions, there was the sense that marginal loss refunds should be related to actual losses in some fashion for some period to reflect those historical decisions, even if not on a transaction basis.

Table 5.12. ISO surplus collection in dollars due to congestion and losses

		Total Surplus
Day-ahead market	Off-peak	1 169
	Intermediate	4 264
	Peak	24 585
	Total	30 018
Real-time market	Off-peak	1 155 (−14)
	Intermediate	23 605 (+19 341)
	Peak	33 398 (+8 813)
(change from day-ahead)	Total (net)	+28 140

(in Section 5.3.6). Changes in the total surplus between day-ahead and real-time are shown
T12 in Table 5.12.

Turning first to the day-ahead off-peak case shown in Fig. 5.2a, the prices at buses 1, 2, and 3 are \$15.00/MWh, \$16.13/MWh, and \$17.31/MWh, respectively. If the slack bus is assumed to be bus 1, then the LMP price components (using the CAISO, 2005 methodology) are an energy component of \$15/MWh; loss components of \$0.00, \$1.13, and \$2.31 per MWh, respectively, at the three buses; and zero congestion components. The loss component at, for instance, bus 2 is calculated by incrementing the load by 1 MW; to meet that load, 1.0873 more MW of generation is needed from the slack. The value of the 0.0757 MW of losses, evaluated at the slack's LMP of \$15/MWh, is \$1.13/MWh. As shown in Section 5.3.6, the difference between what the ISO collects from demand and what it owes to generators is an auction surplus, which in this case is \$1169 (Table 5.12).

In the day-ahead intermediate scenario shown in Fig. 5.2b, there are both loss and congestion surpluses due to the binding transmission constraint. The prices at buses 1, 2, and 3 are \$15.00/MWh, \$20.00/MWh, and \$20.36/MWh, respectively. Using bus 1 as the slack bus for the purpose of calculating LMP components, the energy component of the price is \$15.00/MWh, while the loss (congestion) components are \$0 (\$0), \$1.18 (\$4.82), and \$3.19 (\$1.17) per MWh, respectively at the three buses. As discussed in Section 5.3.6, the total surplus collected by the ISO is \$4264 (Table 5.12). Using the above LMP components, the congestion surplus portion of this total surplus is \$2163. This means that the loss component is \$2,101, equaling the sum of the loss LMP components times the net withdrawals (\$3857) minus the energy LMP component times the net losses (\$1756).

~~However,~~ this division is arbitrary. If the LMP components were instead based on using the load bus (bus 3) as the slack bus (as in the CAISO, 2005 methodology), the estimates of the surpluses results would have been different. Then the energy component would be \$20.36 (the bus 3 price), and the loss components would have been −\$3.52, −\$2.20, and
AU8 \$0.00 per MWh at the three respective buses.[37] The resulting congestion components would also be negative, being −\$1.84, \$1.84, and \$0.00 per MWh, respectively. The congestion surplus would then be calculated as $(-1.84 \times -1227 + 1.84 \times -190)$ or \$1920. Subtracted from the total surplus of \$4264, this yields a loss surplus of \$2355. Thus, the loss and

[37] ~~For example, a 1 MW load increment at bus 1 would be met by decreasing load at the "distributed load slack" (just bus 3) by 0.827 MW; thus, losses would be lowered by 0.173 MW, which at \$20.36/MWh is worth \$3.52.~~

congestion surpluses based on bus 3 being the slack are each about 10% different from the values based on a bus 1 slack, above.

Finally, if the congestion surplus was instead defined based on flowgate shadow prices, it would instead equal $5.39 (the price for the congested flowgate between buses 1 and 2) times 350 MW (the corresponding flow), or $1886; this would imply a total loss surplus of $4264 − $1886 = $2378. These values are close but not identical to the component surpluses resulting from using the distributed load slack (bus 3) as the slack bus. As noted, this method is not used by ISOs because it does not disaggregate to the bus level.

LMPs are calculated both day-ahead and in real-time, and the value of the ISO surpluses may change between the two markets, as shown in Table 5.12. For example, in the off-peak scenarios shown in Figs 5.2a and 5.3a, since there is no congestion, all the surplus collected by the ISO is due to losses. There is less demand off-peak in real-time than day-ahead, so the ISO effectively "owes back" $14 in loss surplus that was assigned to day-ahead buyers after the resettlement of the sellers' and buyers' real-time positions (as discussed in Section 5.5.5). In practice, the ISO surpluses are not disbursed or refunded on an hourly basis, but are aggregated for later disbursal.

In the intermediate and peak scenarios, there are also congestion surpluses due to the binding transmission constraint. In the day-ahead intermediate market scenario, the congestion and loss surpluses are fairly close in value. However, in the real-time intermediate market scenario, due to the increase in demand that requires the dispatch of the more expensive generator at bus 3 and continued congestion causing an increase in the price differentials between buses 1 and 3 (but also creating a negative differential between buses 2 and 3), the total surplus greatly increases relative to the day-ahead surplus. The ISO thus collects additional surpluses in real-time, which are recovered from real-time buyers. A similar result is seen in the peak scenario.

The interpretation of these results must be done carefully. If market participants hold financial transmission rights and are concerned that the settlements of such rights day-ahead will not collect sufficient revenues compared to congestion charges in real-time, then they can use virtual bids and offers to effectively shift their congestion hedge to real-time. This was discussed in Section 5.3.1. With regard to changes in marginal loss surpluses between day-ahead and real-time, these surpluses in total are refunded to market participants, regardless of which market they occurred in. Some ISOs have specific rules to account for differences in magnitudes between the day-ahead and real-time loss surplus so as to ensure that these differences do not create inefficient scheduling incentives when the loss refunds are determined.

5.7.5. Comparison of PJM and New York ISO market rules

As with many areas of auction market design, there are some differences in the rules in PJM and New York regarding dispersal of ISO congestion and loss charge surpluses. Beginning with congestion surpluses, in both markets, these are used initially to pay the set of awarded financial transmission rights. Any residual congestion surplus is then subsequently refunded to market participants. As of this writing, in PJM, the existing transmission rights are paid in full only if the ISO collected sufficient congestion revenues to do so; otherwise, payments are pro-rated.[38] At the end of each month and over the

[38] This rule is due to change in 2008, at which time PJM will pay all transmission rights in full regardless of shortfalls in congestion revenues, through an uplift charge.

course of the period covered by annual allocations of financial transmission rights, PJM then distributes any residual congestion surplus in five stages (PJM, 2007b, p. 46). In the first stage, any surplus is allocated to holders of financial transmission rights that were deficient with respect to their target allocation for the month. In the second stage, any remaining surplus is allocated to financial transmission right payment deficiencies in prior months of the year. Third, any remaining surplus (after stages one and two) is carried forward to the subsequent month and distributed in the same fashion. Fourth, at the end of the year, any remaining surplus is used to compensate for any deficiency in payments to auction revenue rights. Finally, any remaining annual surplus is distributed to entities that had paid PJM transmission access charges that year in proportion to their demand charges or MW reserved capacity for transmission service into, out of, or through the transmission system.

New York ISO has a different method for settlement of financial transmission rights in the event of a congestion rent shortfall, and hence the rules for disposing of any residual congestion surplus after cashing out the transmission rights are also different. Unlike PJM, in the event of a shortfall in congestion revenues, payments to the financial transmission rights are made wholly by transmission owners in New York via a pass-through to their retail rates (sometimes called "full funding" of the rights). Because of this rule, any residual congestion charge surplus is disbursed to transmission owners.

With respect to losses, ~~as noted above, PJM has not yet included marginal loss components in its LMPs, hence it has not had to distribute marginal loss surplus charges. Instead, it has used a version of average loss charges that do not result in any excess~~ ~~AU9~~ ~~collection by the ISO. In 2007,~~ PJM ~~will~~ implement a marginal loss calculation in locational marginal pricing. ~~In contrast,~~ New York ISO ~~calculates marginal loss charges and hence has rules for disposition of the marginal loss~~ surplus, which it calls the "residual loss payment~~." This surplus~~, along with surpluses collected in other ways (with the exception of congestion cost surplus), is credited against what transmission customers are billed as aggregate ISO costs, which include operational costs and the costs of implementing certain ISO programs, such as demand response payments (NYISO, 2001a, Rate Schedule 1).

5.8. Market Power Monitoring and Mitigation

The basic definitions and procedures of market power monitoring and mitigation were introduced in Section 5.2. Here more detail is provided, in particular on the ISO auction rules for economic withholding. ISO markets typically encompass sufficient geographic territory and enough suppliers that generation market concentration is relatively low by most market-wide measures, except in two cases: first, when transmission constraints bind causing ~~market~~ concentration to rise in sub-regions, and second, when supply is scarce and the lack of demand elasticity allows even small suppliers to raise prices. Both of these market conditions are typically intermittent, and their frequency is a factor in creating the mitigation rules. For example, if a transmission constraint causes a persistent, concentrated sub-market to emerge, sometimes called a "load pocket," then a specific remedy may be required for that sub-market, e.g., requiring some generators to sign cost-based contracts.

5.8.1. Identifying and mitigating exercise of market power

In the US ISOs, three actions are typically identified as the exercise of market power: physical withholding, economic withholding, and uneconomic production. Physical withholding refers to any withdrawal of physical generation capacity from the market, including both units that are self-scheduled and those offered into the auction, that could be construed as an attempt to raise the market price. This could result from unscheduled outages that cannot be justified *ex post-*, changes in a generator's physical parameters in its offer that cause it to reduce its output, and failure to respond to an ISO's dispatch instructions. Forced outages and scheduled maintenance are obviously not considered physical withholding. Physical withholding is typically simple to detect, but can be difficult to prove as intentional exercise of market power in the absence of clear evidence. Hence, ISOs have adopted certain measures, such as a percentage share of a generator's available capability withheld or percentage deviation from a dispatch instruction, before physical withholding is identified as a market rule violation. Mitigation measures are *ex post*, and typically involve either revising the generator's offer or applying penalties.

Economic withholding refers to raising offer prices substantially above marginal cost, including opportunity cost, so as to affect the market clearing price. Typically, an offer that is economically withheld is one that has raised its price sufficiently to not get picked by the auction. However, in electricity markets, short-term demand is largely inelastic – i.e., not price sensitive – in which case there will be times when all offers are picked and any price can theoretically clear the market. One situation in which this could be the case when demand is inelastic is if one or more suppliers are "pivotal," meaning that if they withdraw all their available capability the market could not clear.

Most ISOs have implemented a threefold approach to mitigation of economic withholding in the energy market. The first component is an absolute offer cap on day-ahead and real-time energy, which currently stands at $400/MWh in California and $1000/MWh in the eastern US ISOs. No offer can exceed this cap, but a locational market price can be higher than the cap due to network effects.

A second component, present in some ISO markets but not in others, is offer caps outside the transmission constrained area that are below the absolute cap. In some ISOs, such as New York ISO as described later, these have been implemented in a two-step fashion sometimes called a "conduct-impact test." In the first step (the conduct test), an offer is screened to determined whether it has violated a percentage increase, or absolute dollar amount increase, from a "reference" accepted offer price. The reference offer is typically an average of prior accepted offers and is understood to be a proxy for a competitive market offer. In the second step (the impact test), the ISO determines whether offers that violated the first step also increased the LMPs by some additional threshold. An offer that violates both steps is mitigated to its reference offer price and the auction is re-run.

The third component is offer caps for persistently transmission constrained locations, where presumably generation market power is more prevalent. These tend to be more restrictive than the offer caps for locations that are less frequently transmission constrained. They are also strict caps rather than the screens discussed above.

ISOs have adopted different approaches to monitoring and mitigating the price offers for start-up and no-load. In some ISOs, these are restricted to being modified infrequently, while in other ISOs, where they can change daily, they can be subject to similar offer caps to energy. There are also offer caps in the ancillary services markets; in some cases, where there is administrative scarcity pricing for operating reserves, those prices effectively ~~cap~~ the reserves market price and through simultaneous co-optimization of energy and reserves also set the energy LMPs.

5.8.2. Contractual remedies

There are typically a number of generators in ISO markets that have ~~extreme~~ locational market power because they are needed to operate for reliability purposes or persistently

Els UK Job: CDI ch05-i047172 25-10-2007 7:33p.m. Page:234 Trim:165×240MM Float:Top/Bot TS: Integra, India

to resolve congestion. Some such generators were built to provide transmission support. Others are older generators that are only intermittently used for peak hours but which cannot recover fixed costs through the market. For such units, which are sometimes called "reliability must run" units, a contractual solution may be required. Such contracts may be cost-based or pay the higher of a market price or a contract price.

5.8.3. Comparison of PJM and New York ISO market rules

PJM and New York ISO have a number of differences in their market power monitoring and mitigation rules, stemming from their regulatory histories and aspects of their subsequent market development. The most prominent difference is that PJM uses mandatory cost-based offers as a basis for mitigating transmission constrained supply offers (if needed), while New York ISO has a quite different approach under which the system is subdivided into pre-specified "constrained areas" in which there are ~~moving~~ formulas for offer caps that depend on the frequency of congestion and areas outside the constrained areas, where a conduct-impact test is applied. These differences will be explained below.

PJM began its centralized auction market for energy in April 1998 and operated it for a full year with all supply offers capped at marginal costs.[39] When the bid-based energy market was begun in April 1999, PJM continued this method of mitigating supply offers when an energy offer was from a generator's whose output was altered due to transmission congestion, called an "out-of-merit" generator. Subsequently to beginning its markets, PJM also tightened the rules for start-up and no-load. Currently, a generator can choose between cost-based and price-based offers for these offer components. If it chooses cost-based offers, these can be adjusted daily. However, if it chooses price-based offers, these can only be adjusted twice a year, during enrollment periods (a generator can also switch between cost- and price-based offers in these enrollment periods).

Other rules have addressed physical offer parameters, which can provide a generator with market power with additional means to affect market clearing prices. For example, in PJM, the market monitor found in the summer of 1999 that during peak demand hours when the ISO requested all generators to produce on an emergency basis (called "maximum generation emergency alert" in PJM), certain generators anticipated the shortage and increased their minimum run times to the full day. This allowed them to get paid their offer price (including start-up) in hours of the day when the market price had subsided below their offer. The market rule change to address this issue required that a generator's total offer, including payments for start-up and no-load, could not exceed $1000/MWh during the specific hours of the emergency. In any other hours that the generator would operate due to its minimum run time, it would be a price-taker in the market and would not be eligible for additional revenue sufficiency payments (FERC, 2000).

Sometimes, market power rules are overly restrictive. In PJM, for generators subject to offer caps due to being out-of-merit, concern grew that particularly for persistently offer-capped units, energy market prices were not sufficient for such units to recover long-term variable costs. In PJM, this lead to a reduced application of offer caps through

[39] This was because FERC rejected PJM's first market proposal in 1997 which did not provide sufficient information about generation market power to satisfy regulatory requirements. However, while PJM undertook its market power analysis, FERC allowed it to start market operations in April 1998 with the condition that market power was mitigated in the interim using offers capped at marginal cost. The market with liberalized supply offers began one year later.

Fonts: Palatino & Helvetica 9/11 Margins:Top:4PC Gutter:5PC T.W:30PC open recto 1 Color 49 Lines

two measures. First, PJM developed a new test, called the "three pivotal supplier test," to determine when a transmission constraint was creating a truly uncompetitive market behind the constraint. The offer cap is not applied to a generator that affects a particular constrained transmission path if there are three or fewer suppliers that are jointly pivotal with respect to the constraint and if the owner of the generator when combined with the two largest other suppliers affecting the constraint is not pivotal (PJM Tariff, Section 5.6.4). Second, PJM allows generators that are frequently mitigated to recover a greater percentage over and above their incremental cost offer cap.[40]

PJM has also established market power mitigation rules for its regulation and reserves markets. At the start of the regulation market in 2000, PJM expressed some concern about the concentration ~~of the regulation market~~ (i.e., the low number of potential sellers) and as a result was allowed to establish the requirement that regulation price offers cannot vary by hour and later also that they would be subject to a $100/MWh cap. Moreover, due to market concentration some generators in PJM can only submit cost-based offers for regulation. In the spinning reserve market, several zones have caps based on cost-based rates.

New York ISO initially began its markets in November 1999 with generation offer restrictions only within New York City, an obvious load pocket, called a "constrained area." FERC approved this approach based on New York ISO's demonstration that the market outside the constrained area was sufficiently un-concentrated. However, in 2000, New York ISO and its external market monitor developed the "conduct-impact" approach to screening supply offers described above, which has subsequently been adopted directly or in modified form by the other ISO markets, with the exception of PJM. As a component of this method, New York developed a market-based approach to measuring a benchmark competitive offer (and hence to avoid requiring generators to submit marginal cost offers, as in PJM). The basic method is to use the average of a generator's accepted offers in the prior 90 days as a reference price. If there is not a sufficient history of accepted offers, then there are other methods for setting reference prices. The offer caps under this rule are shown in Table 5.13. As in PJM, the rules differ between constrained areas and T13
those in largely unconstrained areas. This is not surprising since a market-based reference offer would reflect the market concentration in the constrained area. Hence, as shown in Table 5.13, the New York ISO resorts to offer caps that are a declining function of the average price in the constrained areas and the number of hours that the area is congested.

Like PJM, New York ISO has also had some experience with overly restrictive mitigation rules. Generators in New York City was initially subject almost continuously to offer caps set roughly at the level of their variable production costs. In 2004, these rules were revised to conform to the conduct-impact approach used in other parts of the system (Potomac Economics, 2005). This reduced ~~to~~ frequency of mitigation and allowed prices to rise to levels more reflective of scarcity within the area.

5.9. Other Topics in ISO Market Design and Implementation

This section reviews several other topics in ISO market design and implementation that have been prominent in the United States in the period under review. The first of these

[40] Specifically, for units that are offer capped for (a) between 60% and 70% of their run hours, the offer cap is either incremental cost plus 10% or incremental cost plus $20/MWh; (b) between 70% and 80% of their run hours, the offer cap is either incremental cost plus 15% or incremental cost plus $30/MWh, and (c) 80% or more of their run hours, the offer cap is either incremental cost plus 10%, incremental cost plus $40/MWh, or unit-specific going forward costs in agreement with the ISO.

Table 5.13. Thresholds for identifying conduct that may lead to economic withholding in New York ISO

Offer Components	Outside Constrained Area	Within Constrained Area (excluding New York City)
Energy and minimum generation offers: ~~d~~Day-ahead market	Lower of 300% or \$100/MWh increase; Offers below \$25/MWh excluded.	When constraint binds, lower of (a) thresholds for outside Constrained Area, or (b) a threshold calculated as follows: (2% × Average Price × 8760)/Constrained Hours, where Average Price is the average day-ahead market price in the Constrained Area over the prior 12 months (adjusted for fuel price changes) and Constrained Hours is the total number of hours in the day-ahead market over the prior 12 months in which any transmission interface or facility leading into the Constrained Area where the generator is located had a shadow price > 0 in any interval.
Energy and minimum generation offers: real-time market	Lower of 300% or \$100/MWh increase; Offers below \$25/MWh excluded.	Same as day-ahead calculation, with the Average Price and Constrained Hours calculation being made using real-time market data. The Average Price is also adjusted for out-of-merit generation dispatch as feasible and appropriate.
Start-up price offer	Increase of 200%.	Increase of 50%.
Regulation and operating reserves offers	Lower of 300% or \$50/MWh increase; offers below \$5/MWh excluded.	Lower of 300% or \$50/MWh increase; offers below \$5/MWh excluded.
Time-based offer parameters (including start-up times, minimum run times, minimum down times)	Increase of 3 hours or increase of 6 hours for multiple time-based bid parameters.	Increase of 3 hours or increase of 6 hours for multiple time-based bid parameters.
Offer parameters in units other than time or dollars (including ramp rates and maximum stops)	Increase of 100% for parameters that are minimum values; 50% decrease for parameters that are maximum values.	Increase of 100% for parameters that are minimum values; 50% decrease for parameters that are maximum values.

Source: New York ISO Market Services Tariff, Attachment H. Available at: www.nyiso.com

is the interaction between the daily energy auction markets and longer-term markets and other functions undertaken by the ISO. The second is the continued existence of market "seams" that include boundaries of market operations that do not conform to natural operational boundaries as well as differences in market design and other factors that may result in economic inefficiency. The third is the important role of software design and development in expanding the scope of the auction markets and capturing opportunities for efficiency.

5.9.1. Longer-term ISO markets and operational/planning functions

In addition to the energy and ancillary service markets described in this chapter, which take place on a daily and hourly basis, the ISO also operates markets and undertakes reliability, operational, and planning functions that take place on longer time-frames, from several days to several years. With respect to markets, these include most notably auctions for generation installed capacity and financial transmission property rights. Table 5.14 lists several of these market and other functions and their respective time-frames. The design of these longer-term ISO markets is discussed in several chapters in this volume and will not be reviewed here. Notably, none of the ISOs currently operate forward energy auction markets on a longer time-frame than the day-ahead period discussed in this chapter. T14

These longer-term markets and functions have interactions with the daily markets, both intended and unintended. For example, the capacity product is the operable MW of a generator and is sold separately from forward or spot energy. In the ISO markets, there is no need to purchase energy and capacity from the same generator. However, most of the capacity products are designed implicitly or explicitly as options for the ISO to call on energy from a generator in the event of shortages. If a capacity generator is called on for energy by the ISO, then it must curtail any sales outside the ISO market and provide real-time energy to the zone for which it has been designated as a capacity resource.[41] As ISOs have found out, it is important to design capacity markets that have sufficient incentives or penalties to enforce this call option. Another type of possible linkage, this

Table 5.14. Longer-term ISO markets and operational/planning functions

Time-frame	Market Functions and Operations	Reliability and Planning Functions
Multi-year	Forward markets for installed capacity Allocation and auction of multi-year financial transmission rights	Regional planning and expansion (transmission, generation, and demand response)
Annual, monthly	Allocation and auction of annual, seasonal and monthly financial transmission rights	Co-ordination of planned transmission and generation maintenance
Weekly	Scheduling of generation with more than one-day start-up times	Load forecasting

[41] Hence, under current designs, a capacity contract is not equivalent to a forward energy contract – i.e., physically linked to a particular buyer of power – because the ISOs are not equipped to implement a priority ranking among wholesale buyers in the event of a load curtailment. Instead, the capacity product is defined zonally, as an option to deliver power from a generator to a zone. For example, PJM conducts a "deliverability test" for generators requesting to be capacity resources using a power flow model that evaluates how the power from that generator will diffuse in a region of the market. When transmission constraints limit the flow of the generator in this test, it must pay for upgrades to fully qualify as a capacity resource. Once it meets the eligibility requirements, a capacity resource is then invested with energy market obligations in exchange for capacity payments, as noted earlier.

time unintended, between forward and spot markets is between holdings of financial transmission rights and energy market behavior. A market participant that both owns generation and can collect congestion revenues from its transmission rights may have the incentive to alter its output so as to maximize the revenues from both energy and its transmission right revenues. This would result in an inefficient dispatch. Such activity has been observed in the markets, but is generally rare.

5.9.2. Market seams

A market "seam" refers to differences between methods of power system operations, market designs, and rules for crossing market boundaries (i.e., as an importer or exporter of power) that create transactions costs or externalities across the boundary. In the United States, the ISO markets can have seams with other ISO markets or with the regions that do not have organized markets. In each case, the primary issue across the seam is transmission scheduling, congestion management, and unscheduled flows (such as "loop flows").

FERC has sought to reduce or eliminate these market seams over the years. In late 1999, it issued a rule that encouraged and provided incentives for all utilities in the United States to join large Regional Transmission Organizations (RTOs) on a voluntary basis (FERC, 1999). A subsequent FERC initiative to merge PJM, New York, and New England on the basis that they already shared a similar market organization failed in this period due to regional differences. In 2002, with RTO formation lagging, FERC proposed a "standard market design" for all utilities in the country that largely followed the basic ISO design described in this chapter (FERC, 2002). However, this proposal met strong political and industry resistance in some regions of the country and was formally revoked in 2005. Hence, for the period covered by this chapter, market seams, both between ISO markets and between the ISO markets and the purely bilateral markets, have remained salient concerns from both an operational and economic perspective. Had the standard market design rule been implemented, many of these seams problems would have been resolved.

5.9.3. Developments in market software

In practice, software has been a limiting factor in the development of efficient market designs. The existing ISO software and data systems are a result of market start-up decisions as well as patches resulting from continual change and improvement. Consequently, changes to a single software system may require changes to many software and data systems. Currently, there is a significant backlog of improvements in each ISO. In part the backlog is due to the extensive testing and changes necessary to install new software modules. Hence, there are still significant efficiencies to be gained from the standardization of data systems so that when improved software is developed in one area it can be easily tested and put into production elsewhere.

Nevertheless, in some areas, ISOs and commercial vendors have been able to introduce innovations that demonstrate the value of their investment in software and growing technological expertise. These innovations have produced economic benefits. One example is the search for faster solution times and closer to optimal solutions for the unit commitment auctions that take place day-ahead and then over the operating day. In the past, the Lagrangian relaxation solution algorithm used by utilities and then by the ISOs was known to reach sub-optimal solutions, but the size of the problems that it could solve and its solution times had been reduced over the years. In theory, mixed integer programs could achieve an optimal solution but had difficulties with solution speed for large problems. In recent years, these solution time issues have been reduced such that ISOs can now deploy mixed integer programs for the auction markets (Johnson et al., 1997; Hobbs et al., 2001; Guan et al., 2003). In addition, mixed integer programs allows a much more detailed and full specification of the unit commitment problem and relieves the market co-ordinator of many of the simplifying assumptions that are necessary using Lagrangian relaxation.[42]

These advantages and the tractability of mixed integer programming algorithms have led several ISOs to introduce or test mixed integer program-based implementations over Lagrangian relaxation. Due to the computational complexity of unit commitment problems, ISOs which implement mixed integer program based algorithms tend not to solve their unit commitment problems to optimality due to limitations on solution times.[43] Streiffert et al. (2005) note that the enhanced modeling capabilities of mixed integer program allows the ISO to deal directly with a number of constraints that were very difficult to model in Lagrangian relaxation. PJM has implemented mixed integer programming for its day-ahead market with estimated savings of $54 million per year due to efficiency improvements in the auction solution. While savings from the more precise solutions provided by mixed integer programs may be a small percentage of total generation costs, i.e., 1–4%, a 1% savings in generation costs translates into a $1–$2 billion annual cost saving in the United States alone.

5.10. Extensions of the Market Design

The US ISOs have been operating spot energy auctions with locational marginal pricing since 1998, and regulation and operating reserves markets with various designs for almost as long. As discussed in this chapter, market designs and software are continuously being refined in response to various factors, including incomplete markets, the introduction of new technologies and software, and also due to the greater penetration of some existing technologies that create new operational requirements, such as demand response and wind energy. The auction designs described in this chapter should be adaptable, such that the basic organizational and pricing principles will remain appropriate as technological changes take place.

Among the near-term design challenges relevant to this chapter is more “complete” pricing of ancillary services, such as reactive power. Currently, many systems are dispatched without using a full AC optimal power flow, and impose overly restrictive voltage

[42] Even if the mixed integer program algorithm times out before finding an optimum, one is still left with a primal-feasible solution and a bound on the optimality gap. These intermediate solutions are often found within the same amount of time a Lagrangian relaxation-based algorithm takes, and typically have optimality gaps of the same size or smaller than Lagrangian relaxation commitments. Furthermore, a mixed integer program-based solution algorithm allows ISOs to easily introduce new types of unit-operating and system constraints to the formulation of the problem, whereas Lagrangian relaxation-based techniques generally require extensive reprogramming of the feasibility heuristics to ensure that the unit commitment satisfies all the necessary conditions.

[43] PJM, for instance, allows its mixed integer program optimizer to run within a certain period of time or until the optimality gap is below some maximal threshold and uses whatever intermediate integer-feasible solution the solver has found. If an ISO is left to rely on an intermediate integer-feasible but sub-optimal solution, the same issues of generator payoffs, energy pricing, and inequity of the resulting dispatch arise as with sub-optimal Lagrangian relaxation commitments.

levels. Reactive power is not priced or is priced inappropriately. The design question is whether ISOs should sign long-term contracts for reactive power that include obligations to perform or whether spot markets for reactive power can elicit a more efficient outcome and overcome concerns about locational market power (Hogan, 1993; Kahn and Baldick, 1994; FERC, 2005). To implement such markets, optimal power flow software needs to be improved and integrated with unit commitment models.

Another line of inquiry concerns the more active participation of "dispatchable" transmission elements in the energy spot markets (O'Neill et al., 2005). That is, with sufficient regulatory oversight, both merchant and regulated transmission owners could offer some or all of their transmission capacity into the ISO markets at a price. Another aspect of this development could be the unit commitment of transmission elements; e.g., the ability to decouple a transmission line if that leads to a reduction in the auction objective function.

5.11. Conclusions

The US ISO markets, many encompassing multiple states and reaching a geographic scope that was not anticipated just a few years earlier, are a major achievement of electricity regulatory reform. The auction market designs for energy and ancillary services have developed on the basis of both theoretical principles and practical decisions. In some cases, design mistakes were made, but equally the ISOs and their stakeholders have worked to refine the designs and learn from experience, with FERC oversight. This has led to a high degree of convergence on key elements of the designs, such as day-ahead and real-time markets with locational marginal pricing, the reliability unit commitment, and the co-optimization of regulation and reserves; although as this chapter has shown, there remain many design differences between the ISO markets.

A lesson of the first decade of the ISO markets is that an efficient spot market for electric power that respects economic principles and reliability requirements ends up being rather complicated. There are many products and their pricing and financial settlement rules are often difficult for market participants to understand and analyze. This chapter has sought to provide a step-by-step review of many of these market rules and procedures and to explain why design choices were made. More recently, concerns about the cost-benefit ratio of implementing such markets have been raised. Among other things, this has prompted calls for simplification of the market designs. Although simplification where possible (and increased standardization) of market rules in the United States should be a design objective, observers should also note that the complicated design of the daily auction markets is in part because so much of the power system has been exposed to transparent market pricing and procedures. What in prior years were system operational decisions whose costs were internalized by utilities, often inefficiently, are now integrated into daily market auctions and generally priced efficiently. Transmission usage is priced on the margin and optimized over large regions, resulting in much more efficient use of transmission capacity. Extensions of these locational pricing principles and the application of unit commitment to transmission elements could yet result in even further gains in utilization of the existing grid. However, as discussed, there are several aspects of market pricing that do continue to rely for practical reasons on types of average pricing. Most notably, most ISOs still charge buyers load-weighted average LMPs on a zonal basis rather than the LMPs at their nodes.

The current market designs will continue to evolve with refinement of the market rules, changes in technology, and shifting regulatory requirements. The markets will function better if technology and consumer interest (or regulation) allow a robust demand response to emerge. At least some of the regulatory and reliability aspects of the market designs, such as supply offer caps and reliability unit commitments, could become less important if that takes place.

Finally, this chapter has focused on the short-term, daily markets. As discussed in other chapters of this book, there is still evolution in the design of other elements of electricity market design to support investment decisions, such as resource adequacy or capacity markets and long-term financial transmission rights. Because of the interactions between these different design elements, all aspects of the ISO market designs must be carefully integrated and calibrated [see, e.g., Stoft (2002) and O'Neill et al. (2006) for discussion].

~~5.12.~~ 5A. Appendix: Mathematical Formulation of the Auction Examples

The auction examples given in several sections of this chapter employ a simplified version of the ISO auction designs, but with many of the main features of those markets, including realistic network power flows including congestion and losses and including both start-up and energy offers by generators. In this appendix, the mathematical formulation of the auction model used in the examples is given, with some additional explanation. There are different ways to write this auction mathematically; the approach taken here is intended to improve the intuition of the model. For a more detailed mathematical description of the ISO auction on transmission networks, see, e.g., O'Neill et al. (2002) and the ISO auction manuals.

A single period version of the auction model is as follows.

$$Max \sum_{m,i}(\mathrm{BID}_{mi}d_{mi}) - \sum_{m,i,h}(\mathrm{START}_{mih}z_{mih} + \mathrm{OFFER}_{mih}g_{mih}),$$

$$\text{subject to } \sum_{m,h} g_{mih} - y_i = \sum_m d_{mi}, \quad \forall i, \ (\pi_i) \qquad (1)$$

PLEASE ITALICIZE THE HIGHLIGHTED SUBSCRIPTS ON THIS AND NEXT 2 PAGES

$$f_k^+ \le F_k^{+\max}, \quad \forall k, \quad (\theta_k^+) \qquad (2)$$

$$f_k^- \le F_k^{-\max}, \quad \forall k, \quad (\theta_k^-) \qquad (3)$$

$$-y_i + \sum_k [D_{ik}(f_k^+ - f_k^-) + f_k^- L_{ki}^- f_k^- + f_k^+ L_{ki}^+ f_k^+] \le 0, \quad \forall i, \qquad (4)$$

$$R(f^+ - f^-) = 0, \qquad (5)$$

$$g_{mih} \le G_{mih} z_{mih}, \quad \forall m, i, h, \qquad (6)$$

$$d_{mi}, g_{mih}, f_k^+, f_k^- \ge 0,$$

$$z_{mih} \in \{0, 1\},$$

where the notation is defined as follows.

ADD SPACE

Index sets

H is the set of generators and virtual offers, $h = 1, \ldots, n[H]$.
I is the set of buses, $i = 1, \ldots, n[I]$, in the transmission system.
K is the set of transmission facilities, $k = 1, \ldots, n[K]$.
M is the set of participants in the auction market, $m = 1, \ldots, n[M]$, whether as sellers or buyers (physical and virtual).

Variables

d_{mi} is the quantity of energy bought by market participant *m* at bus *i*.

f_k^+, f_k^- represent the flow on transmission element k in the positive and negative direction respectively (defined arbitrarily). The net real power flow on transmission element k could thus be defined as f_k, where $f_k = (f_k^+ - f_k^-)$. f^+, f^- are the vectors of these flows.

g_{mih} is the quantity of energy, physical, or virtual, sold by market participant m at bus i from generator or virtual offer h.

y_i is the amount of real power injected at node i (withdrawn at node i if $y_i < 0$) that is induced by the d bids and g offers that were awarded through the auction.

z_{mih} is the commitment decision associated with the offer for generator h submitted by market participant m at bus i. It is either 0 (uncommitted) or 1 (committed). In a dynamic model, separate variables would be defined for the start-up decision (1 signaling a start-up occurs in a given hour) and commitment (1 indicating that the generator is operating in that period).

$\pi_i, \theta_k^+, \theta_k^-$ are Lagrange multipliers associated with selected sets of primal constraints in the auction.

Parameters and operators

BID_{mi} is the bid price (\$/MWh) submitted by market participant m at node i associated with a demand quantity d_{mi}. In this context, "bid" means a bid to buy energy.

OFFER_{mih} is the offer price (\$/MWh) submitted by market participant m for generator or virtual supply h at node i associated with quantity g_{mih}. In this context, "offer" means a offer to sell energy.

START_{mih} is the start-up price (\$) submitted by market participant m for generator h at node i. For all physical generators, this quantity is non-negative. Note that for all virtual offers, it is exactly zero.

D is the arc incidence matrix, $\{D_{ki}\}$. $D_{ki} = 1$ if $(f_k^+ - f_k^-)$ represents a MW flow out of bus i through transmission line k in a positive direction; $D_{ki} = -1$ if the flow through k is in a negative direction; and $D_{ki} = 0$ otherwise.

$F_k^{+\max}, F_k^{-\max}$ are transmission capacity constraints – thermal, stability, or contingency limits – associated with a transmission element k in the positive and negative directions.

G_{mih} is the upper bound on the capacity offered by market participant m for generator or virtual supply offer h at node i.

L_{ki}^-, L_{ki}^+ represent resistance loss coefficients (decrease in imports to bus i) due to a negative and positive flow, respectively, through transmission line k.

$R = \{r_{vk}\}$ are line reactances used in Kirchhoff's Voltage Law analogues. r_{vk} is the value of reactance for transmission line k that appears in voltage loop v. $r_{vk} = +R_k$ or $-R_k$ if line k occurs in loop v, depending on whether a positive $(f_k^+ - f_k^-)$ is in the same or opposite sense of flow around v. $r_{vk} = 0$ if link k does not occur in loop v. Consistent with the linearized DC model of load flow (Schweppe et al., 1988), the number of independent loops v must be equal to $K - N + 1$, where K is the number of lines considered and N is the number of buses.

The objective function maximizes social welfare, defined as the sum of consumer surplus and producer surplus. This is the same as the integral of the demand curve (sum of accepted demand bids) minus as-bid production costs. Production costs include commitment costs, which are incurred if a generator is operating (i.e., $z_{mih} = 1$), along with variable generation costs. More general versions include start-up and min run costs as separate terms in the objective. Constraint (1) is a net energy balance requirement for each bus on the network, whose dual variable is the LMP. Equations (2) and (3) are transmission capacity constraints on each transmission element. Their dual variables are the so-called "flowgate prices." Equations (4) and (5) are DC analogues to Kirchhoff's Current and Voltage Laws. Equation (6) is the generation upper operating limit; if the unit is not committed (i.e., $z_{mih} = 0$), then this constraint forces MW generation to be zero. Lower operating limits ("min run constraints") and ramp rate constraints are not shown, but could be introduced.

The numerical example takes place on the three-bus network in Fig. 5.1, in which the arrows show the direction of flow for an injection at bus 1 and a withdrawal at bus 3 (note that the arrows do not correspond to the direction of the flowgates). In the equations that follow, the three transmission lines indexed k above are labeled for the two buses to which they are connected in the "positive" direction. Hence, the line from bus 1 to bus 2 is labeled "12," the line from bus 1 to 3 is labeled "13," and the line from bus 2 to 3 is labeled "23." All loss factors on all lines = 0.00001 [MW/MW^2]. All reactances, $R_k = 1$. Then (4) for each bus becomes

$$\begin{aligned}
\mathrm{KCL}_1 &: -y_1 + (f_{12}^+ - f_{12}^-) + (f_{13}^+ - f_{13}^-) + (0.0001 f_{12}^-)^2 + (0.0001 f_{13}^-)^2 \le 0,\\
\mathrm{KCL}_2 &: -y_2 - (f_{12}^+ - f_{12}^-) + (f_{23}^+ - f_{23}^-) + (0.0001 f_{12}^+)^2 + (0.0001 f_{23}^-)^2 \le 0,\\
\mathrm{KCL}_3 &: -y_3 - (f_{23}^+ - f_{23}^-) - (f_{13}^+ - f_{13}^-) + (0.0001 f_{23}^-)^2 + (0.0001 f_{13}^-)^2 \le 0,
\end{aligned}$$

and (5) becomes

$$\mathrm{KVL} : (f_{12}^+ - f_{12}^-) + (f_{23}^+ - f_{23}^-) - (f_{13}^+ - f_{13}^-) = 0.$$

With these simplifications, the auction examples can be replicated using commercially available software, such as GAMS.

Acknowledgment

The authors would like to thank David Mead, Partha Malvadkar, and Harry Singh for useful comments, and Perry Sioshansi for editorial advice and great patience.

References[44]

~~Ausubel, L. and Cramton, P. (1996).Demand reduction and inefficiency in multi-unit auctions. Working Paper No. 96-07, Department of Economics, University of Maryland.~~ ~~AU10~~

California ISO (2005).Locational Marginal Pricing (LMP) Study 3C, Analysis of Market-Based Price Differentials, Description of Methodology, Appendix A, Disaggregation of LMP Components. 19 November. Available at http://www.caiso.com/14cd/14cd6d735cf60.pdf.

Cramton, P. and Stoft, S. (2006). Uniform-Price auctions in electricity markets. Working Paper, University of Maryland, 20 March.

~~Elmaghraby, W., O'Neill, R.P., Rothkopf, M.H., and Stewart, W.R., Jr. (2004). Pricing and efficiency in 'lumpy' energy markets. *Elec. J.*, **17**(5), 54–64.~~ ~~AU11~~

[44] The technical manuals referenced below were all current as this chapter was being written. Such manuals are periodically updated to reflect market rule changes. The websites of the ISOs whose materials are cited are as follows: PJM, www.pjm.com; New York ISO, www.nyiso.org; ISO-New England, www.iso-ne.com; Midwest ISO, www.midwest.org

Fabra, N., von der Fehr, N., and Harbord, D. (2004). Designing electricity auctions. Center for the Study of Energy Markets, University of California, Berkeley, CSEM WP-122, February .

Federal Energy Regulatory Commission (FERC) (1996a).*Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities*. Order No. 888, FERC Stats. & Regs. 31,036, 24 April.

Federal Energy Regulatory Commission (FERC)(1996b).*Open Access Same-Time Information System (formerly Real-Time Information Networks) and Standards of Conduct*. Order No. 889, 75 FERC 61,078,

AU12 24 April.

AU13 Federal Energy Regulatory Commission (FERC) (1997). PJM Interconnection, L.L.C., et al., *Order Conditionally Accepting Open Access Transmission Tariff and Power Pool Agreements, Conditionally Authorizing Establishment of An Independent System Operator and Disposition of Control over Jurisdictional Facilities, and Denying Rehearing*. 81 FERC 61,257, 25 November.

Federal Energy Regulatory Commission (FERC) (1999a). Central Hudson Gas & Electric Corpora-

AU14 tion et al., *Order Conditionally Accepting Tariff and Market Rules, Approving Market-Based Rates, and Establishing Hearing and Settlement Judge Procedures*, 86 FERC 61,062, 27 January.

Federal Energy Regulatory Commission (FERC) (1999b).*Regional Transmission Organizations*. Order No. 2000, 89 FERC 61,285, 20 December.

AU15 Federal Energy Regulatory Commission (FERC) (2000). PJM Interconnection, L.L.C., *Order Accepting Tariff* [illegible]013, 7 July.

Add reference Harvey, S.M., Hogan, W.W. and Pope, S. (1997). For rest of of reference see our written comments.

Federal [illegible]n (FERC) (2002).*Standard Market Design and Structure, Notice of Propos*[illegible] Docket Number RM01-12-000, September.

Federal [illegible]n (FERC) (2005). *Principles for Efficient and Reliable Reactive Power Supply*[illegible]Report, AD05-1-000, 4 February.

Guan, X. [illegible], A. (2003). Optimization based methods for unit commitment: Lagra[illegible] mixed integer programming. *Power Engineering Society General Meetin*[illegible]

AU16 Helman, [illegible]itoring and mitigation in the US wholesale power markets. *Energy*[illegible]

Hobbs B[illegible]d O'Neill, R.P.(2000).Evaluation of a truthful revelation auction in the context of energy markets with nonconcave benefits. *J. Reg. Econ*., **18**(1), 5–32.

Hobbs, B.F., Stewart, W.R. Jr., Bixby, R.E., et al. (2001). Why this book?: New capabilities and new needs for unit commitment modeling. In *The Next Generation of Electric Power Unit Commitment Models* (B.F. Hobbs, M.H. Rothkopf, R.P. O'Neill, and H-P. Chao, eds). Boston: Kluwer Academic Press, pp. 1–14.

Hogan, W.W. (1993). Markets in real electric networks require reactive prices. *Energ. J*., **14**(3), 171–200.

Hogan, W.W. (2006). Resource sufficiency guarantees and cost allocation. Comments submitted to the Federal Energy Regulatory Commission, Docket No. ER04-691-065, 25 May . Available at http://ksghome.harvard.edu/˜whogan/.

AU17 ~~Hogan, W.W. and Ring, B.J. (2003). On minimum-uplift pricing for electricity markets. John F. Kennedy School of Government, Harvard University, Cambridge, MA, March. Available at http://ksghome.harvard.edu/˜whogan/.~~

Johnson, R.B., Oren, S.S., and Svoboda, A.J. (1997).Equity and efficiency of unit commitment in competitive electricity markets. *Util. Pol*., **6**, 9.

Kahn, A.P., Cramton, P., Porter, R., and Tabors, R. (2001). Pricing in the California Power Exchange electricity market: Should California switch from uniform pricing to pay-as-bid pricing? Blue Ribbon Panel Report, Prepared for the California Power Exchange, January.

Kahn, E. and Baldick, R. (1994). Reactive power is a cheap constraint. *Energ.J*., **15**(4), 191–202.

AU18 ~~Kamat, R. and Oren, S.S. (2002). Rational buyer meets rational seller: Reserves market equilibrium under alternative auction designs. *J. of Reg. Econ*., **21**(3), 247–88.~~

Klemperer, P. (1999). Auction theory: A guide to the literature. *J. of Econ. Surv*., **13**(3), 227–86.

Krishna, V.(2002). *Auction Theory*. San Diego, Calif.: Academic Press.

~~Midwest ISO (2005a). Midwest Markets Initiative, *Business Practices Manual for Energy Markets*, Manual No. 002, Version 4, Updated 15 March.~~

~~Midwest ISO (2005b). Midwest Markets Initiative, *Business Practices Manual for Energy Markets Instruments*, Manual No. 003, Version 5, Updated 25 April.~~ ~~AU19~~

New York ISO (1999). *Transmission and Dispatching Operations Manual*, Manual 12, September.

New York ISO (2001a). *Market Services Tariff*. Issued on 16 January.

New York ISO (2001b). *Day-Ahead Scheduling Manual*, Manual 11, June. AU20

New York ISO (2004). *Subject: Scheduling a "Must-Run" Generator*, Technical Bulletin 026, 10 November.

New York ISO (2005). *Subject: Allocation of Uplift Costs to Load and Other Entities Associated with Virtual Trading*, Technical Bulletin 082, 5 November, 2001; updated 4 November, 2005.

New York ISO (2006). *Ancillary Services Manual*, Manual 2, May.

O'Neill, R. P., Baldick, R., Helman, U., Rothkopf, M.H., and Stewart, W.R.(2005). Dispatchable transmission in RTO markets. *IEEE Trans. on Pow. Sys.*, **20**(1), 171–9.

O'Neill, R. P. and Helman, U. (2007). Regulatory reform of the U.S. wholesale electricity markets. In *Creating Competitive Markets: The Politics of Regulatory Reform* (M.K. Landy, M.A. Levin, and M. Shapiro, eds). Brookings Press. AU21

O'Neill, R., Helman, U., Hobbs, B.F., and Baldick, R. (2006). Independent system operators in the USA: History, lessons learned, and prospects. In *Electricity Market Reform: An International Perspective* (F. Sioshansi and W. Pfaffenberger, eds). Amsterdam: Elsevier.

O'Neill, R.P., Helman, U., Hobbs, B.F., et al. (2002). A joint energy and transmission rights auction: proposal and properties. *IEEE Trans. Pow. Sys.*, **17**(4), 1058–67.

O'Neill, R.P., Sotkiewicz, P.M., Hobbs, B.F., et al. (2005). Efficient market-clearing prices in markets with nonconvexities. *Eur. J. Op. Res.*, **164**, 269–85.

PJM (2000). *State of the Market Report*, *1999*. Market Monitoring Unit, PJM Interconnection, L.L.C., June.

PJM (2001). *State of the Market Report*, *2000*. Market Monitoring Unit, PJM Interconnection, L.L.C., June.

PJM (2002). *State of the Market Report*, *2001*. Market Monitoring Unit, PJM Interconnection, L.L.C., June.

PJM (2003). *State of the Market Report*, *2002*. Market Monitoring Unit, PJM Interconnection, L.L.C., 5 March .

PJM (2004). *State of the Market Report*, *2003*. Market Monitoring Unit, PJM Interconnection, L.L.C., 4 March.

PJM (2005).*State of the Market Report*, *2004*. Market Monitoring Unit, PJM Interconnection, L.L.C., 8 March.

PJM (2006a).*State of the Market Report*, *2005*. Market Monitoring Unit, PJM Interconnection, L.L.C., 8 March.

PJM (2006b)*PJM Manual 15: Cost Development Guidelines*. PJM Interconnection, L.L.C., Revision 07, Effective Date: 3 August.

PJM (2007a).*State of the Market Report*, *2006*. Market Monitoring Unit, PJM Interconnection, L.L.C., 8 March.

PJM (2007b). *PJM Manual 11: Scheduling Operations*. PJM Interconnection, L.L.C., Revision 30, Effective Date: 20 March.

PJM (2007c).*PJM Manual 6: Financial Transmission Rights*. PJM Interconnection, L.L.C., Revision 9, Effective Date: 4 April.

PJM (2007d). *PJM Manual 12: Balancing Operations*. PJM Interconnection, L.L.C., Revision 15, Effective Date: 15 May.

PJM News (2004). PJM'S New Generation-Scheduling Software to Save Customers Estimated $56 Million: Unit Commitment System Uses First of its Kind Technique. Valley Forge, Pa., 24 June.

Potomac Economics, Ltd. (2003). *2002 State of the Market Report*, *New York ISO*. Independent Market Advisor to the New York ISO, June.

Potomac Economics, Ltd. (2004). *2003 State of the Market Report*, *New York ISO*. Independent Market Advisor to the New York ISO, May.

Potomac Economics, Ltd. (2005). *2004 State of the Market Report*, *New York ISO*. Independent Market Advisor to the New York ISO, July.

Potomac Economics, Ltd. (2006). *2005 State of the Market Report*, *New York ISO*. Independent Market Advisor to the New York ISO, August.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., and Bohn, R.E. (1988). *Spot Pricing of Electricity*. Boston: Kluwer Academic Publishers.

Stoft, S. (2002). *Power System Economics*. New York: IEEE Press.

Streiffert, D., Philbrick, R., and Ott, A. (2005). A mixed integer programming solution for market clearing and reliability analysis.I In Power Engineering Society General Meeting, 2005, IEEE, San Francisco, CA.

Sweeney, J.L. (2006). California electricity restructuring, the crisis and its aftermath. In *Electricity Market Reform: An International Perspective* (F. Sioshansi and W. Pfaffenberger, eds). Amserdam: Elsevier.

Wilson, R. (2002). Architecture of power markets. *Econometrica*, **70**(4), 1299–340.

Chapter No: 5

Query No	Contents
AU1	Please Query to PUBLISHER: First and third author have same affiliation. Please advise how to present affiliations.
AU2	Please Please provide citation for reference “FERC, 1999”.
AU3	Please Reference citaion “Stoft and Cramton, 2006” has been changed to “Cramton and Stoft, 2006” as per the reference list. Please check is this OK?
AU4	Please ‘from New York’ has been changed to ‘in New York’. Is this OK?
AU5	Please Reference citation “PJM, 2006” has been changed to “PJM, 2006a” as per the reference list. Please check is this OK?
AU6	Please ‘Heretofore...reflect both the effect of marginal....’ has been changed to ‘Heretofore...reflect the effect of both the marginal....’. Is this OK?
AU7	Please Reference “Hogan et al,. 1997” has been cited in the footnote ‘33’ but not provided in the list. Please check.
AU8	Please note 17 and note 37 are similar. Is this OK?
AU9	Please provide an update.
AU10	Please provide citation for Reference “Ausubel and Cramton, 1996”.
AU11	Please provide citation for Reference “Elmaghraby et al. (2004)”.
AU12	Please provide citation for Rference “FERC (1997)”. Please check.
AU13	Please Is it a part of the title?
AU14	Please Is it a part of the title?
AU15	Please Is it a part of the title?
AU16	Please provide citation for Reference “Helman (2006)”
AU17	Please provide citation for Reference “Hogan and Ring (2003)”.
AU18	Please provide citation for Reference “Kamat and Oren (2002)”.
AU19	Please provide citations for the References “Midwest ISO (2005a, b)”.
AU20	Please provide citation for reference “New York ISO (2001b)”.
AU21	Please provide citation for Reference “O’Neill and Helman (2007)”.