

INTEGRATED ELECTRICITY MARKETS IN NEW YORK: DAY-AHEAD AND REAL-TIME MARKETS FOR ENERGY, ANCILLARY SERVICES, AND TRANSMISSION

Bradley Kranz and Robert Pike
New York Independent System Operator
Altamont, New York

Eric Hirst
Consulting in Electric-Industry Restructuring
Oak Ridge, Tennessee

November 2002

1. INTRODUCTION

Electricity has at least two unique characteristics that complicate real-time operations and competitive wholesale markets:

- The need for continuous and near instantaneous balancing of generation and load, consistent with transmission-network constraints; this requires metering, computing, telecommunications, and control equipment to (1) monitor loads, generation, and the transmission system; and (2) adjust generation output to match load.
- The transmission network is primarily passive, with few “control valves” or “booster pumps” to regulate electrical flows on individual lines; control actions are limited primarily to adjusting generation output and to opening and closing switches to add or remove transmission lines from service.*

The primary mechanism a system operator uses to meet both responsibilities is the dispatch of generation (or dispatchable load).# Because the system operator in a restructured electricity industry owns no generation, it must obtain these generation and load resources from market participants through offers and bids submitted in centralized markets.

Real-time energy markets and operations are identical and cannot be separated because of the technical complexities of managing the grid and the speed of real-time actions (Chandley

*In addition, the transmission grid is surprisingly fragile when pushed to its limits. Unlike highways or telephone systems, overloading transmission elements can cause considerable equipment damage and lead to sudden, major outages (Stoft 2002).

#To a lesser extent, the system operator can change transformer settings, capacitor banks, and other devices to adjust voltages or switch transmission lines in or out of service.

2001). Only a single system operator can do this. Also, the system operator is the only entity that has all the information needed to make real-time decisions. This information includes data on the current status of each generating unit (e.g., real- and reactive-power outputs, ramp rate, and maximum output) and each transmission element (e.g., flows, voltages, and phase angles). This information is crucial to making decisions that maintain security* of the transmission system.

Although not essential for reliability, it is logical for the system operator to also run day-ahead markets for energy, ancillary services, and congestion management. These short-term forward markets parallel the real-time markets the system operator *must* run. It is logical because only the system operator can ensure that (1) the financially-binding contracts struck in day-ahead markets are physically feasible (i.e., they violate no transmission or other reliability limits) and (2) are consistent with the results of the real-time markets.# In addition, the results of the day-ahead markets are crucial inputs for the real-time markets and operations. Splitting the markets between two entities would increase transaction costs for market participants, the forward market manager, and the real-time operator.

Finally, bilateral contracts and other long-term arrangements made through decentralized arrangements are tied to the day-ahead and real-time markets, through these spot prices. That is, buyers and sellers of long-term power seek prices consistent with what they expect the spot prices to be (Fig. 1).

If these spot markets are competitive and efficient, they provide substantial benefits. They ensure that the demand for electricity is met at the lowest possible production cost, consistent with

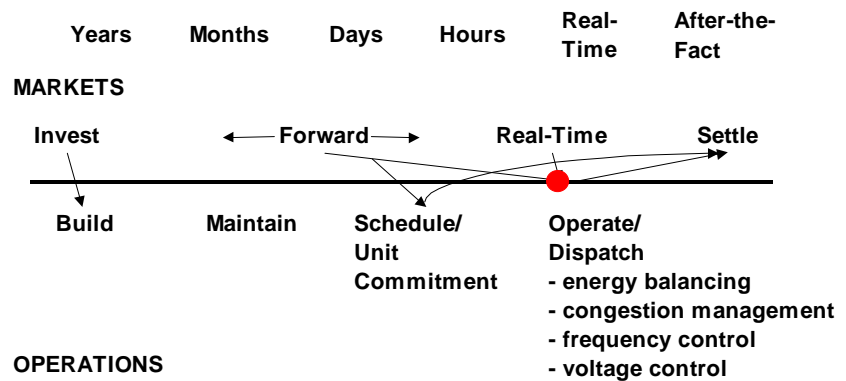


Fig. 1. Relationship between markets and operations as a function of time.

*The definition of reliability used by the North American Electric Reliability Council (NERC 2002) encompasses two concepts, *adequacy* and *security*. Adequacy is defined as “the ability of the system to supply the aggregate electric power and energy requirements of the consumers at all times.” NERC defines security as “the ability of the system to withstand sudden disturbances.” In plain language, adequacy implies that there are sufficient generation and transmission resources available to meet projected needs plus reserves for contingencies. Security implies that the system will remain intact even after outages or other equipment failures occur. One of the system operator’s primary responsibilities is to maintain security.

#This consistency means that if nothing changes between day ahead and real time, the prices and quantities settled in the day-ahead markets will match those in real time.

security constraints. They ensure that consumption goes to those customers that most highly value electricity. And they ensure that, at each time and location, the right amount of power is produced and consumed at the correct price.

This paper explains the day-ahead and real-time markets the New York Independent System Operator (NYISO) operates; the major changes underway to improve the efficiency of these markets; how these markets compare with the Standard Market Design (SMD) proposed by the U.S. Federal Energy Regulatory Commission (FERC); and how the New York markets compare with those of the other independent system operators (ISOs) in the Northeastern United States and Ontario, Canada.

2. HISTORY AND EVOLUTION OF NEW YORK MARKETS

The NYISO began operations in November 1999. It is responsible for ensuring the reliability of the New York power grid, which includes compliance with the planning and operating standards issued by NERC and the Northeast Power Coordinating Council. In addition, the ISO administers the FERC-approved transmission tariff and the associated market rules.

The ISO grew out of the New York Power Pool, which had controlled the New York transmission system and the real-time dispatch of the generating units in the state since the 1960s. Although the Power Pool was owned by the member utilities in New York (seven investor-owned utilities and the New York Power Authority), the ISO is independent of all market interests, including the owners of generation and transmission.

The New York electricity market serves almost 19 million people with a peak load of almost 31,000 MW. The New York electricity infrastructure includes more than 330 generating units with a total installed capacity of 36,000 MW, and almost 11,000 miles of transmission lines. A defining, perhaps unique, feature of the New York electrical system is the severity of its transmission constraints. These constraints often limit the flow of power from the north and west into New York City and Long Island. The value of the energy, transmission services, and ancillary services* that flow through the NYISO markets amounts to about \$5.5 billion a year. The “raw” energy accounts for 72% of wholesale energy costs in New York, with congestion and losses accounting for another 19% (Fig. 2).

The New York markets have been, from the opening day, comprehensive and sophisticated. That is, the ISO runs integrated markets for energy, congestion, losses, and

*The NYISO administers markets for four real-power ancillary services: regulation, 10-minute spinning reserve, 10-minute reserve (which includes spinning and nonspinning reserves), and 30-minute reserve (which includes the 10-minute reserves). In addition, the ISO provides system control and dispatch and acquires, through nonmarket mechanisms, the generation resources needed for voltage control and system blackstart.

ancillary services, both day ahead and real time. In addition, the ISO calculates advisory prices hour ahead. Because these markets are so comprehensive and tightly integrated, various problems occurred during the early months of market operations, especially during the spring of 2000 (New York State Electric & Gas Corp. 2000; FERC 2000).

The most important of these problems relates to inconsistencies among the models the ISO used to dispatch generation. For the day-ahead market and security analysis, the ISO uses its Security-Constrained Unit Commitment (SCUC) model to calculate the

physically feasible flows and associated market-clearing prices. The ISO's Balancing Market Evaluation (BME) produces hour-ahead advisory prices, adjusts interchange schedules on the basis of these prices, and decides whether to commit any quick-start generators (units that can start within 10 or 30 minutes). Finally, the ISO uses its Security-Constrained Dispatch (SCD) model to perform an economic (least-cost) dispatch of the available generating units every five minutes. Although the SCUC and BME models contain the same algorithm, the SCD model uses a different approach because it is based on a decades old legacy system. As a consequence, the hour-ahead and real-time prices sometimes diverged, especially when shortages occurred.

Among the inconsistencies between BME and SCD is the treatment of reserves. BME sets aside capacity to fully meet the expected energy and reserve requirements for the three reserve services, but SCD can use the 30-minute reserves for energy in real time. This different treatment of 30-minute reserves can lead to lower real-time prices than those forecast by BME. BME schedules generators to meet peak demand during an hour, while SCD meets the load during each 5-minute interval, which suggests that BME overcommits units. This overcommitment can also lead to lower prices in real time.

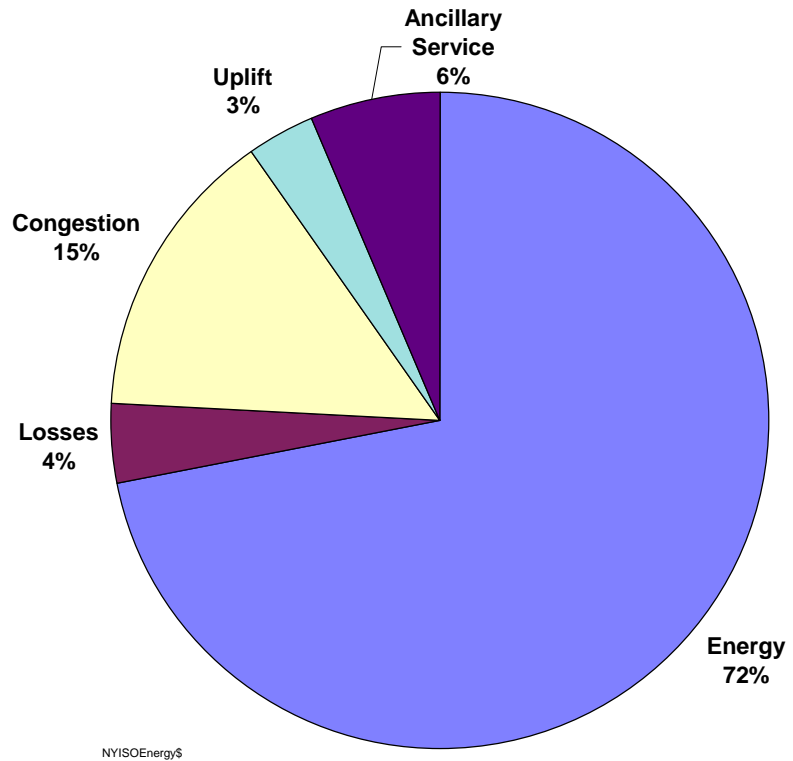


Fig. 2. The components of New York's \$1.7 billion wholesale energy cost for summer 2002 (Patton 2002).

BME uses forecasts of generator and transmission status produced 75 minutes ahead of real time. Because of the lag between forecasts and real time, these inputs to BME may turn out to be incorrect. SCD uses current information on the actual status of generation and transmission equipment. In addition, BME can not handle physical interchange schedules. Instead, it manages these schedules on the basis of economics (differences in energy prices inside and outside New York).

Since market opening, the ISO, working with its market participants, has made many changes to its systems to improve the accuracy of price signals and otherwise enhance the efficiency of its markets. The Independent Market Advisor to the ISO noted that the markets in 2001 were “workably competitive, with limited instances of significant withholding or other strategic conduct” (Patton and Wander 2002).

Some of the improvements include the introduction of virtual trading in November 2001, the treatment of recallable exports as 30-minute reserves under certain conditions, better alignment of market rules with the neighboring ISOs to permit more interchange schedules and fewer problems with these schedules being inappropriately cut by BME, the initiation of demand-response programs, and the recognition of latent reserves within BME. Also, as of October 2001, the New York ISO became the first North American entity with locational markets (both quantities and prices) for reserve services. The amounts of reserves and the resulting prices can differ across three zones (West, East, and Long Island).

Figure 3 illustrates the current scheduling and operations used by the New York ISO (1999a and b). The Bid/Post system allows market participants to post generator and load bids, as well as schedules for bilateral transactions. The system is also used by the ISO to post results from the day-ahead and real-time markets, as well as the advisory results of the hour-ahead BME.

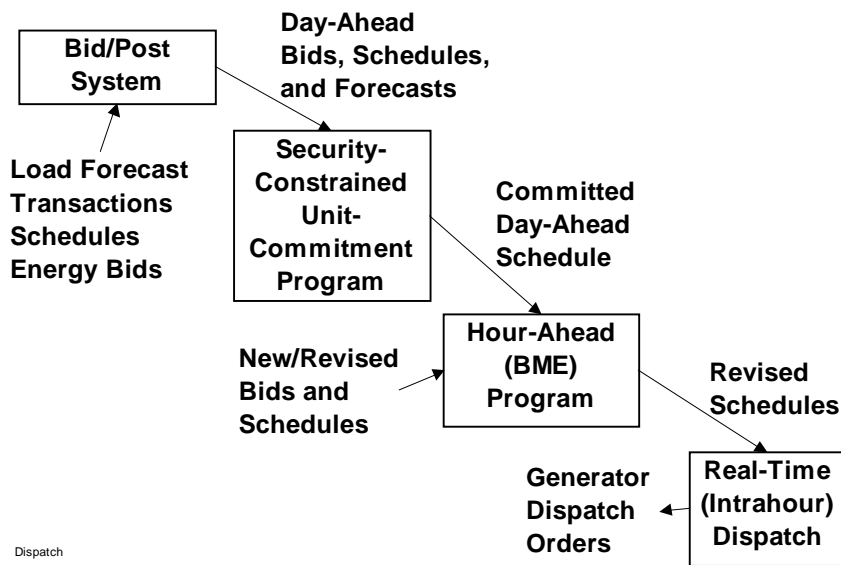


Fig. 3. New York ISO process for day-ahead scheduling and real-time operations.

SCUC

The objective of the SCUC is to minimize the total bid production cost of meeting all purchasers’ bids to buy energy a day ahead, provide enough ancillary services, commit enough

capacity to meet the ISO's load forecast, and meet all bilateral transaction schedules (including imports and exports) submitted to the ISO.

The ISO's day-ahead scheduling process includes assembling data on planned transmission outages and the ISO's zonal load forecasts as inputs to the SCUC model. The SCUC considers several factors, including "current generating unit operating status, constraints on the minimum up and down time of the generators, generation and start up bid prices, unit-related startup and shutdown constraints, minimum and maximum generation constraints, generation and reserve requirements, maintenance and derating schedules, transmission constraints, phase angle regulator settings, and transaction bids" (New York ISO 1999b). Market participants submit bids and offers to the ISO by 5 am, and the ISO publishes SCUC results by 11 am of the day before the operating day.

A complete run of the SCUC program includes five passes (three commitment runs plus two dispatch runs) to get a final solution for the following day's schedule of generation. The commitment runs include: (1) based solely on generator and load bids plus bilateral transactions and ancillary-service requirements, (2) commitment of any additional generation needed if the ISO load forecast is higher than the market-participant schedules, and (3) security-constrained dispatch to ensure that all the first-contingency requirements are met (Exhibit 1). SCUC results include hourly schedules for energy and the four ancillary services as well as hourly locational prices for energy* (calculated for about 400 generator nodes), prices for the reserve services for three zones, and aggregate prices for regulation. The energy prices include the costs of marginal losses and transmission congestion.

Generator inputs to the SCUC include up to six constant-cost segments of an energy curve, as well as the no-load cost (\$/hr), startup bid (\$), and startup and shutdown constraints (the number of times each day a unit can be stopped). The model adds transmission-loss penalties for each generator to account for system losses.

The SCUC results, in terms of scheduled quantities and prices, represent binding financial contracts. Any deviations between schedules and real-time generator outputs or consumer loads are settled in the real-time market. In addition, the ISO guarantees that any generator scheduled by the ISO to run some time during the operating day will, at a minimum, recover its out-of-pocket costs. In other words, if the revenues from the energy market are not enough to cover the unit's startup, no-load, and fuel costs over the full 24-hour period, the ISO will cover these costs, called the bid-production cost guarantee.

*Generators in New York are paid for their output on the basis of nodal prices, while loads pay for power on the basis of zonal prices (New York has 11 zones). Each zone represents an aggregation of nodes with the zonal price equal to the generation-weighted average of the nodal prices.

Exhibit 1. NYISO Security-Constrained Unit Commitment

The five passes include three commitment (scheduling) runs and two dispatch runs in sequence:

1. SCUC solves for bid load, virtual load, and virtual supply, while securing the bulk transmission system.
2. SCUC solves for ISO-forecast load ignoring local reliability rules. All the generators committed in pass 1 are included. Additional units may be committed to meet the forecast load (if higher than the scheduled load).
3. SCUC solves for ISO-forecast load and local reliability rules. Some of the units from pass 2 may be decommitted and replaced with other units to meet local security requirements.
4. SCUC solves for forecast load and local reliability requirements. The units committed in pass 3 are dispatched.
5. SCUC solves for bid load, virtual load, and virtual supply. The units in pass 4 plus the combustion turbines from pass 1 are dispatched to determine the day-ahead market clearing prices.

Source: NYISO (2002a).

BME

About 75 minutes before each hour, the BME unit-commitment model is run, with results posted 45 minutes before the operating hour. BME commits (and decommits) units to meet the requirements for energy, regulation, and reserves. Bids into the BME (used to determine the resources available for real-time dispatch) can include resources that are dispatchable within 5 minutes plus fixed-block energy (nondispatchable) available for the next hour. In addition, BME can evaluate modified and proposed new bilateral transactions, offers to sell energy in the real-time market, and transaction bids in external markets. Market participants can modify their bids and schedules up to 75 minutes before the operating hour.

BME uses ISO forecasts as inputs to its calculation, whereas the inputs to SCD (discussed below) rely primarily on current conditions on the grid (obtained from the ISO's supervisory control and data acquisition system). In addition, BME commits generation to meet the next hour's expected *peak* demand, while SCD dispatches generation to meet the demand

for the next five minutes. These differences between BME and SCD, among others, can lead to substantial differences in the prices estimated by the two models.

SCD

For real-time dispatch, the New York ISO classifies resources as on dispatch, off dispatch (but online), or offline but available. According to NYISO (1999a), “The function of the SCD program is to determine the least-cost dispatch of generation within the NYCA [New York Control Area] to meet its load and net interchange schedule, subject to generation, transmission, operating reserve, and regulation constraints. SCD performs this function nominally every five minutes as part of the real-time operation of the NY Power System.” The SCD objective of minimizing cost is limited to the *incremental* bid cost of generation participating in the spot market.

Typically, less than 5% of the electricity consumed in New York is bought and sold in real time. About half the electricity is obtained through bilateral contracts and the remainder is purchased in the ISO’s day-ahead energy market.

Just as the SCUC is run multiple times, the SCD is run twice, once for feasibility and the second time for optimization. The first run seeks a solution that meets load, the second run tries to improve on the first solution by finding a lower-cost combination of generators without violating any of the real-time security constraints. The New York SCD considers about 200 contingency states. Inputs to the analysis include telemetry values of generation output, power flows on the transmission system, load, net interchange, and calculated losses. Additional SCD runs may be needed to turn combustion turbines (CTs) on or off.

Resources can set the 5-minute market-clearing price only if they are not being dispatched against one of their limits. In addition, resources outside the control area cannot set the market price because of constraints on hourly interchange schedules. The 5-minute prices calculated by SCD are used to settle any differences between the day-ahead schedules and real-time operations, at the nodal level for generation and the zonal level for loads. These prices are *ex ante*, that is, they are set at the beginning of each 5-minute interval based on what the model expects generators and loads to do during the following 5 minutes.

EXAMPLE

The weather on July 29, 2002, was hotter than expected, leading to a peak load of 30,700 MW, almost 6% higher than forecast the day before (Fig.4).

As shown in Fig. 5, energy prices, both day-ahead and real-time were higher in New York City than in the West zone, a reflection of transmission congestion. This congestion

increased average prices in New York City by \$20 to \$30/MWh. During mid-afternoon (when loads were highest), congestion costs increased sharply, to \$50/MWh day ahead and almost \$80/MWh in real time.

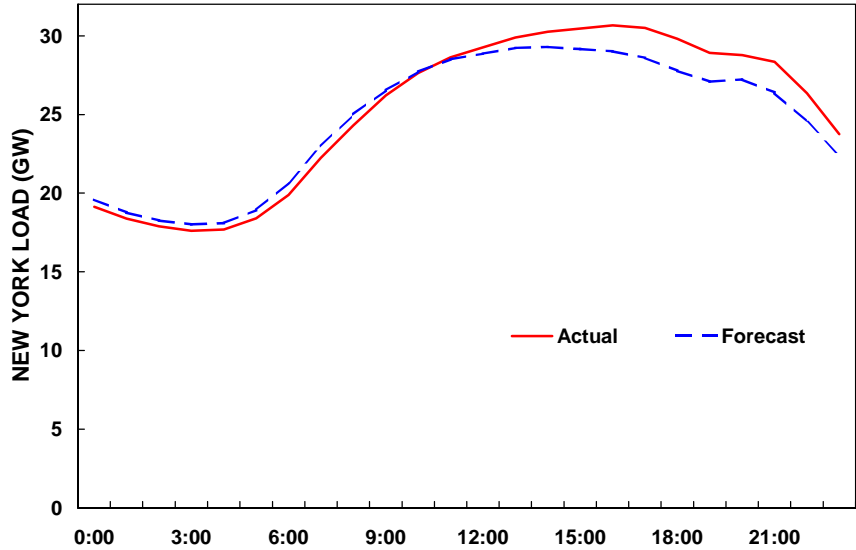


Fig. 4. Actual and forecast load for New York on Monday, July 29, 2002.

Figure 5 also shows how, from 2 to 6 pm, real-time prices were about \$600/MWh above the day-ahead prices. These prices reflect the shortages the ISO was experiencing during that time, which required it to buy several hundred megawatts of emergency purchases and invoke reserve pickup several times that afternoon.

Finally, the prices for ancillary services, especially for regulation, were quite variable that day (Fig. 6).^{*} As expected, the price of regulation is greater than that of the other services, and the price of spinning reserve is higher than those of the other reserve services.[#]

3. PLANNED REAL-TIME SCHEDULING SYSTEM

The New York ISO is currently working on the design and implementation of a successor to BME and SCD. The planned replacement, called the Real-Time Scheduling (RTS) system, consists of two components, Real-Time Commitment (RTC) and Real-Time Dispatch (RTD).[§]

^{*}These prices do not include the unit-specific payments for opportunity costs, which are collected from customers through uplift. The ancillary-service prices calculated in the Real-Time Dispatch model (discussed below) will explicitly include opportunity costs, which will then no longer be collected through uplift.

[#]The unusual pattern of regulation prices could be a consequence of the number and types of generators that bid into this market relative to the energy market hour by hour. Specifically, low-cost hydro units likely provided regulation during the day, while more expensive fossil units provided regulation at night.

[§]In addition, the ISO is adding an up-to-date state estimator to its system. This new state estimator will provide much more detailed and accurate information to RTC and RTD on the current status of the New York grid, which will yield commitment and dispatch requests that are more consistent with the physics of the bulk-power system.

RTS is expected to be fully operational in 2004. Implementation of RTS is expected to reduce uplift costs by more than \$16 million a year. These reductions are a consequence of better price consistency between RTC and RTD, real-time scheduling of ancillary services to minimize operating costs, and reduced use of out-of-merit calls for generation.

Table 1 summarizes the key differences between the current BME/SCD system and the planned RTS system.

REAL-TIME COMMITMENT

RTC performs a security-constrained unit commitment once every 15 minutes (analogous to the once-a-day SCUC). The focus of RTC is on (1) adjustments to interchange schedules (including exports, imports, and

wheels) on the basis of expected prices and (2) decisions on whether to start up or shut down CTs that can start within 10 or 30 minutes. This intraday unit-commitment program will supplement the day-ahead scheduling performed by SCUC, which focuses on the larger steam units that require more time to start up and shut down.

RTC's unit commitment will be conducted more frequently than is true for BME (every 15 minutes instead of every hour), will use more current information (30 minutes ahead of dispatch instead of 75 minutes), will post results closer to real time (15 minutes ahead of commitment decisions instead of 45 minutes ahead), and will have a longer optimization

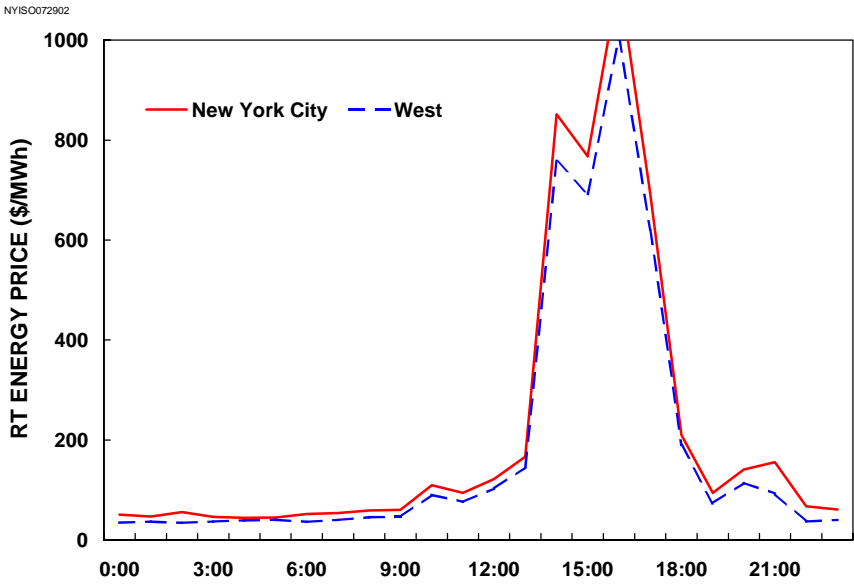
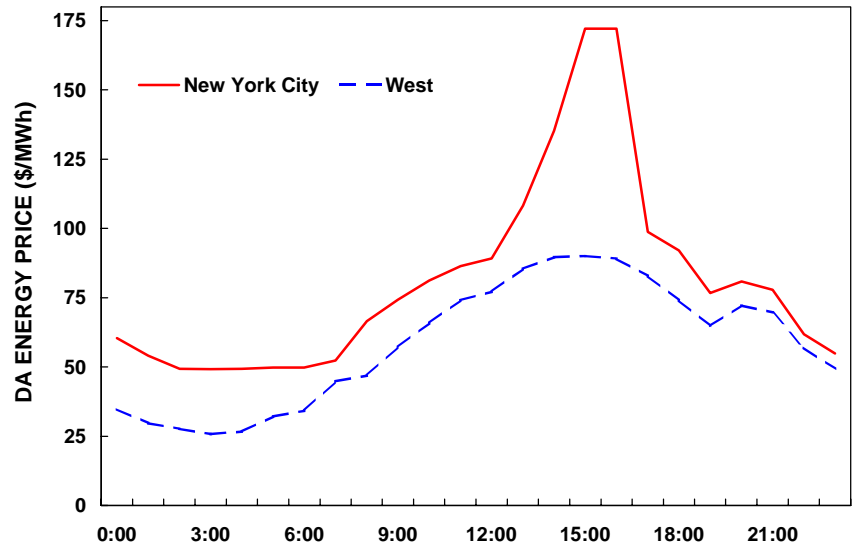


Fig. 5. Day-ahead (top) and real-time (bottom) energy prices in the New York City and West zones for July 29, 2002.

horizon (150 minutes instead of 120 minutes); see Table 2. The unit-commitment process can adjust interchange schedules every 15 minutes, which should permit more efficient exports and imports. RTC will accept new or revised generator offers and demand bids up to 60 minutes before the start of an operating hour.

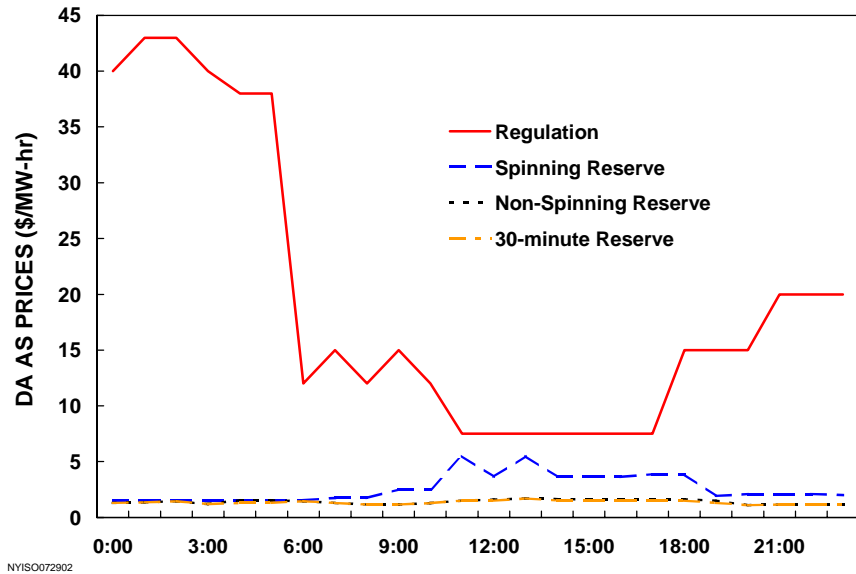


Fig. 6. Day-ahead prices for the real-power ancillary services in New York on July 29, 2002.

RTC will use outputs from a new state estimator, in addition to raw telemetry, to provide more accurate (as well as more timely) inputs to the unit-commitment calculations. Improving the accuracy, internal consistency, and timeliness of the input data on generator and grid conditions should lead to more accurate solutions (i.e., ones that better reflect actual conditions and are truly least cost).

Table 1. Differences between BME/SCD and RTS

BME and SCD	RTS solutions
BME and SCD use different network models	RTC and RTD use the same model
BME selects generation to meet hourly peak demand, while SCD meets 5-minute load	RTC conducts unit commitment over 2-½-hour period, using 15-minute snapshots; RTD dispatches resources every 5 minutes based on 60-minute assessment
BME requires resources to meet load plus all reserve requirements; SCD holds 10-minute reserves but can dispatch 30-minute reserves in merit order as needed	RTC and RTD treat reserves in an identical fashion, using the same price-responsive demand curve
SCD, because it considers only the next 5-minute interval, can yield erratic dispatch price patterns	RTD, because it optimizes the dispatch over 60 minutes, should yield smoother and more realistic patterns
BME commits 30-minute CTs based on a 1-hour analysis; SCD commits 10-minute CTs based on a 5-minute analysis	RTC commits both 10- and 30-minute CTs based on a 2-½-hour analysis; RTD commits no units

Source: DePillis (2002).

Table 2. Comparison of BME/SCD and RTC/RTD time horizons (minutes) ahead of decisions^a

	Initialize input data	Optimization period	Post results
BME	75	120	45
RTC	30	150	15
SCD	0	5	5
RTD	10	60	5

^aRTC can adjust schedules and transactions, both inside the New York control area and with other control areas, every 15 minutes, compared with only once an hour for BME.

The RTC run during the first 15-minute interval of an hour, with results posted at :15 minutes, makes the following types of decisions:

- Decides which 10-minute CTs should start up and be running at their minimum generation level by :30 minutes of the current hour,
- Decides which 30-minute CTs should start up and be running at their minimum generation level by :45 minutes,
- Decides which 10- and 30-minute CTs are to be shut down and disconnected from the grid by :30 minutes,
- Schedules economically bid external transactions for the following hour, and
- Schedules pre-scheduled (physical) external transactions for the following hour.

The remaining three RTC runs for an hour make similar start and stop decisions on CTs but do not modify interchange schedules. Running RTC more frequently and closer to real time will ensure that the decisions to turn on or off CTs and to modify interchange schedules will be more economically efficient. Figure 7 shows the sequence and timing of RTC and RTD runs.

Both RTC and RTD will accommodate price-sensitive load and demand-response programs more readily than the current system does. “Load that meets all metering requirements* and has demonstrated its ability to respond to dispatch instructions can bid into RTC. RTC will commit and dispatch the resource as if it were just another generator on the system. Loads that are able to respond on a shorter time frame than the 15-minute notice provided by RTC may treat the RTC schedules as advisory and wait for dispatch signals from

*These requirements include real-time metering and communications to permit receipt of 5-minute basepoints.

RTD. ... Contractual interruptible loads will be evaluated like dispatchable loads if there are prices associated with the contracts. If there are no prices associated with the interruptible loads they will be handled by the operators in real time ... ” (Hartshorn, Kranz, and de Mello 2002).

In addition, the new system will incorporate demand curves for all reserves. This is an important change because it will reflect the reality that operating reserves are not infinitely valuable (Exhibit 2). Indeed, control-area operators routinely convert reserve capacity into energy rather than disconnect retail loads to maintain reserve margins.

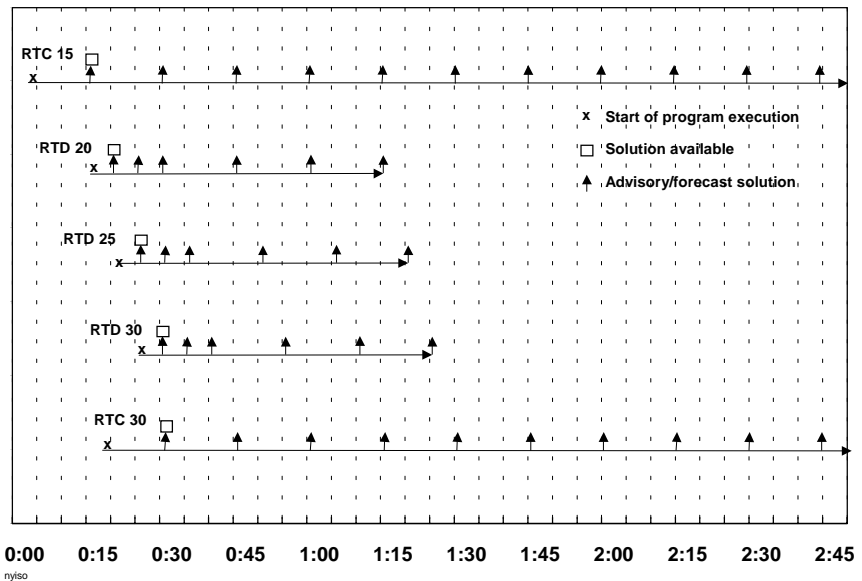


Fig. 7. Timelines for RTC and RTD.

REAL-TIME DISPATCH

RTD dispatches, on a 5-minute basis, the units that are scheduled by SCUC and RTC. “RTD is a multi-period security-constrained dispatch model that co-optimizes to simultaneously solve load, reserves, and regulation. RTD makes no unit-commitment decisions. It simply dispatches the resources available to it on a least-as-bid-cost basis” (Hartshorn, Kranz, and de Mello 2002). RTD makes dispatch decisions on the basis of a full 60 minutes, instead of the 5 minutes currently used by SCD. This longer-term perspective will eliminate some of the intrahour volatility in dispatch signals and prices that can occur with the present SCD. The longer time horizons in RTC and RTD should also reduce the frequency of out-of-market calls and yield lower uplift costs.

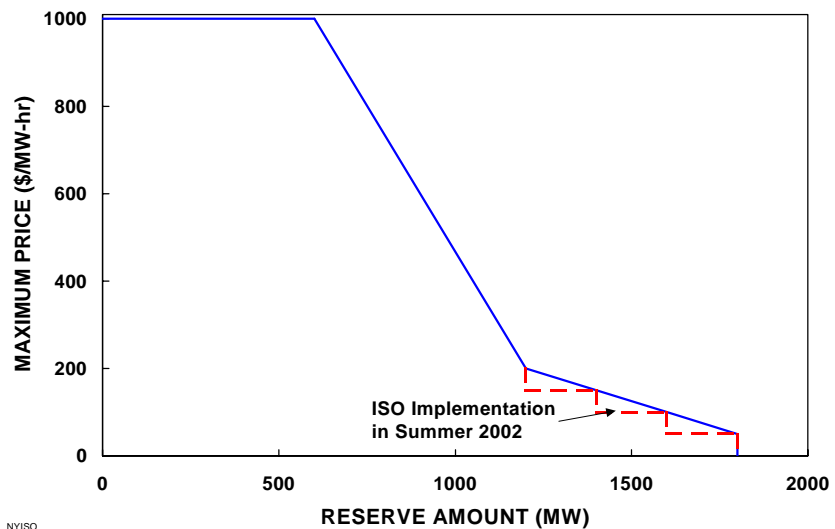
In addition, the energy and ancillary-service prices set by RTD will be ex post, based on the *actual* performance of the units on dispatch during the prior 5-minute interval (rather than ex ante, based on the *expected* performance of these units, as in SCD). The RTD treatment of reserves will be fully consistent with that within RTC. And the ancillary service prices, calculated as the shadow costs from the model solution, will be fully consistent with each other and the energy prices. As a consequence of these and other changes, the consistency of prices from day-ahead through hour-ahead to real-time should be much greater than in the past. This improved consistency, as well as better scheduling of units for ancillary services and prices that more accurately reflect the physical conditions on the grid, will reduce the amount of uplift

Exhibit 2. Demand Curves for Reserves

The NERC and Northeast Power Coordinating Council reliability rules on reserve requirements are deterministic. That is, they specify minimum amounts of spinning reserve, supplemental reserve, and 30-minute reserve as functions of the first and second largest contingencies. The rules imply that these reserve amounts are infinitely valuable, i.e., no price is too high to pay to maintain the required quantities.

These rules are inconsistent with both economic theory and actual operations. Theory suggests that the value of reserves can be no more than the value of lost load. Implicitly, when regulators impose price caps in wholesale markets (\$1000/MWh in New York), they are estimating the value of lost load. Thus, the New York ISO should pay no more than \$1000/MW-hr for reserves. Control-area operators often run short of reserves when the price of additional resources is extremely high or when the only way to maintain reserves is to involuntarily interrupt some load.

RTC and RTD will include demand curves for all reserves, which will recognize that the ISO may be deficient in reserves when resources are scarce. In addition, RTC and RTD will permit gradual restoration of reserves following a reserve pickup (rather than the instantaneous recovery now reflected in BME and SCD). This gradual recovery will prevent post-contingency price spikes from occurring.



Possible demand curve for reserves. The dashed line shows the implicit demand curve used during Summer 2002, based on the ISO's treatment of exports. The ISO requires about 600 MW of spinning reserve, 1200 MW of 10-minute reserves, and 1800 MW of 30-minute reserves.

collected in New York. In other words, the hourly market-clearing prices will more accurately reflect the physical situation (generation, demand, and transmission) and what customers actually pay for electricity.

The optimization across energy and ancillary services ensures that ancillary service prices fully incorporate the opportunity costs associated with provision of energy or the other services (i.e., prices reflect the shadow prices from the optimization model). Also, RTD updates its assignment of reserves every five minutes, which permits the best use of all resources (energy vs reserves). These 5-minute updates to ancillary-service assignments and prices ensure that suppliers are indifferent to providing energy, regulation, or reserves and, thus, are motivated to follow RTD dispatch instructions.

The RTD-Corrective Action Mode (RTD-CAM) “is a specialized version of RTD that will only be run under extraordinary circumstances at the request of the system operators. RTD-CAM will have the capability to commit 10-minute (Fast Start) Gas Turbines. RTD-CAM will be run on demand and produce schedules in about 30 seconds [RTD takes about three minutes to execute] from kickoff” (New York ISO 2002b). RTD-CAM is used to restore the New York electrical system to normal operations after the loss of a major generator or transmission line. A CAM run differs from a normal RTD run in several ways: (1) generator and dispatchable-load basepoints are constrained by emergency, rather than normal, ramp rate and operating limits; (2) a 10-minute, rather than 5-minute, target is used; and (3) capacity set aside as 10-minute reserves is released for energy production. The RTD-CAM pricing mechanism will not artificially suppress energy or reserve prices during emergencies, which can occur when capacity assigned to reserves is released to provide energy. Specifically, RTD-CAM will try to maintain the required levels of 10-minute reserves. Even if sufficient resources are not available during reserve pickups to meet these requirements, the energy and reserve prices will be sufficiently high to reflect the scarcity situation that currently exists.

This continuous optimization also ensures that during a reserve pickup (when reserves are activated) energy prices will not fall, as can sometimes occur if forward assignments of reserves are not modified in real time. During a reserve pickup, the 30-minute demand curve remains unchanged, but the hard constraints on 10-minute reserves are removed from RTD. During reserves restoration, demand curves for the 10-minute reserves gradually call for more reserves over time, reflecting the physical reality that reserves can not be restored instantly. This process for restoring reserves will prevent post-contingency price spikes from occurring in the energy and reserves markets.

4. COMPARISON OF NEW YORK WITH FERC’S STANDARD MARKET DESIGN

FERC’s (2002) SMD calls for a two-settlement system, with day-ahead and real-time markets for energy, congestion management, regulation, spinning reserve, and supplemental reserve (Exhibit 3).^{*} The market design specifies a single market-clearing price at each location

^{*}FERC’s SMD discussed several topics not covered in this paper, including network access transmission service, markets for congestion revenue rights, market monitoring, long-term resource adequacy, and transmission planning and expansion

Exhibit 3. SMD Requirements for Day-Ahead and Real-Time Markets

Day-ahead scheduling and markets

Accommodate self schedules and bilateral transactions (both internal and external)
Offer voluntary unit commitment to (1) simultaneously optimize (minimize production cost) across locational energy and all ancillary services (regulation, spinning reserve, and supplemental reserve) and (2) guarantee recovery of operating costs (bid revenue sufficiency guarantee)

Accept multipart bids from loads and offers from generators

Accept bids for virtual (non-physical) supply and demand

Post hourly locational prices and schedules for energy, as well as marginal losses

Post prices and schedules for ancillary services (ancillary service prices based on availability bids plus SCUC calculation of opportunity cost, and use of reverse cascading)

Prepare load forecast; if forecast is higher than day-ahead schedules, commit additional resources (called replacement reserves) to meet incremental load forecast

Intraday scheduling

Accept new and changed schedules and transactions so long as they are consistent with security limits

Accept new and revised bids for real-time markets

Commit additional resources if needed for reliability

Real-time markets

Perform economic (least-cost) dispatch every five minutes to minimize production costs for locational energy and ancillary services

Post interval and hourly prices

Source: FERC (2002).

that reflects the marginal cost and value of energy. Differences in locational energy prices reflect the cost of congestion between those locations. FERC also specified zonal pricing for 10-minute reserves.

The day-ahead markets establish hourly prices for energy and three ancillary services (regulation and the two 10-minute reserves). These prices and the associated quantities are financially firm and physically feasible. Physically feasible means that the transmission system can enable all the scheduled transactions while remaining within all reliability limits. Financially firm means that any real-time deviations from the day-ahead schedules are settled at real-time prices.

Both the day-ahead and real-time markets are bid based, accept multipart bids for energy, require co-optimization across all products, and offer a voluntary unit-commitment service. Interchange schedules can be either economic (based on the difference between internal and external real-time prices) or physical (with the transaction to flow regardless of price). The markets must be fully open to demand resources as well as supply resources.

The New York markets, especially upon implementation of RTS, fully meet the proposed FERC requirements for SMD. Some of the characteristics embodied in the New York system called for in FERC's SMD include: technology- and fuel-neutral rules; accommodation of demand resources in all markets; voluntary participation in these markets (so that bilateral contracts and self-schedules are permitted); prohibition against a requirement for balanced schedules; integrated (i.e., simultaneously optimized over all products and services) day-ahead markets and scheduling for energy, ancillary services, and congestion; integrated real-time markets for energy, ancillary services, and congestion; mitigation of market power in these markets; few penalties for energy imbalance (in favor of market-based charges and payments for imbalances); the use of locational-marginal pricing to manage congestion; and reduced seams issues (in particular, the plan to permit changes to interchange schedules every 15 minutes instead of only hourly).

5. COMPARISON WITH OTHER ISOS

New York is located in the only part of North America dominated by control area operators that are ISOs. These ISOs include PJM (with a peak load of about 54,000 MW), ISO New England (25,000 MW), and the Ontario Independent Electricity Market Operator (25,000 MW).^{*} The markets for energy, ancillary services, and congestion management in these neighboring ISOs are similar to the ones in New York (Table 3).

Although the markets in the four northeastern ISOs share many general characteristics, they differ in their details:

- Regulation is a 5-minute service in PJM and New York, but in New England it is based on both the 10- and 60-minute response of generators providing the service.[#]

^{*}New York is also interconnected with Hydro Quebec (peak load of about 19,000 MW) through DC ties.

[#]By 5-minute service, we mean that a unit's ramp rate (MW/minute) multiplied by 5 is the maximum amount of regulation it can sell into the New York or PJM markets.

Table 3. Comparison of markets among northeastern ISOs

	New York	PJM	New England		Ontario
			Current	Planned ^b	
Number of markets	Day-ahead, hour-ahead, and real-time	Day-ahead and real-time	Real-time	Day-ahead and real-time	Real-time
Market products	Locational energy plus four ancillary services ^a	Locational energy plus regulation and spinning reserve ^c	Energy plus four ancillary services ^a	Locational energy plus regulation and spinning reserve	Energy plus 10- and 30-minute reserves ^d
Locational reserve pricing	Yes	No	No	No	No
Energy bids	3-part	3-part	3-part	3-part	1-part
Day-ahead unit commitment	Voluntary, central	Voluntary, central	Self	Voluntary, central	Self
Congestion-management pricing	Nodal for generation, zonal for load	Nodal for generation and load	Uplift	Nodal for generation, zonal for load	Uplift
Losses	Marginal	Average	Average	Marginal	Average
Integration of markets ^e	Yes	Partial	Partial	Partial	Partial
Installed capability requirements	Yes	Yes	Yes	Yes	No

^aThe four ancillary services include regulation, 10-minute spinning reserve, 10-minute reserve, and 30-minute reserve.

^bThe new design, adopted largely from PJM, is scheduled to begin operation in March 2003.

^cThe PJM markets for regulation and spinning reserve are real time as of December 2002.

^dOntario purchases regulation through long-term contracts.

^eIntegration means the ISO simultaneously optimizes (minimizes the production cost of) the resources to meet the energy, congestion-management, and ancillary-services requirements.

- New York's day-ahead unit commitment process schedules units for regulation and sets the day-ahead regulation price at the same time it schedules units for energy and reserves and sets the prices for these services. In PJM, the regulation market does not open until after the day-ahead energy market has cleared. And this day-ahead regulation market only sets prices; PJM does not assign units to the regulation service until close to real time.
- In real-time, generators that do not closely follow PJM dispatch signals are paid the market clearing price for any overgeneration or pay the market-clearing price for any undergeneration, but they are not subject to penalties for failure to accurately follow dispatch signals. New York, New England, and Ontario, on the other hand, include penalties for units whose output does not closely correspond to the system operator's requests.*

Thus, even though the systems are conceptually similar among the northeastern ISOs, lots of small differences occur in their implementation.

The markets in other parts of North America are, in many ways, different from those in the northeast. For example, the wholesale electricity markets in the Electric Reliability Council of Texas (accounting for about 85% of the Texas load), California, and Alberta do not include day-ahead markets. The forward markets for energy are all decentralized and dominated by bilateral arrangements. In all three areas, the markets for ancillary services are separate from (not coordinated with) the real-time energy market. And in all three areas, congestion is either socialized (with its costs collected through an uplift charge paid by all customers) or zonal (California). Finally, none of these three regions has an installed capability requirement. On the other hand, all three regions have real-time markets for balancing energy and day-ahead (or longer in Alberta) markets for the regulation and reserve services.

6. CONCLUSIONS

Electricity production and consumption must occur at essentially the same time. Therefore, real-time (minute-to-minute) operations and the associated markets and prices are essential ingredients of a competitive wholesale electricity industry. These hourly and intrahour markets are the foundation of all forward markets and contracts, including day-ahead markets, monthly futures, and bilateral contracts. These intrahour operations maintain system reliability by ensuring that enough and the right kinds of supply and demand resources are available when needed at the correct locations (Hirst 2001).

*The NYISO needs these penalties because of the many strict security limits on the New York grid. In PJM, on the other hand, when generators do not closely follow PJM dispatch instructions, PJM adjusts the real-time energy price (e.g., up if generators are dragging) to encourage generators to follow dispatch signals.

New York is installing a state-of-the-art system to ensure that its electricity markets are efficient and competitive. These markets should provide the correct incentives to generator owners to operate their units in a way that is fully consistent with system needs, provide the correct incentives to investors to build new generation and demand-response programs where and when needed, and provide the correct incentives to consumers to use electricity only when the cost of doing so is less than the value. The New York system is probably more advanced than any other now in operation or planned in North America.

New York's planned improvements will address several critical issues concerning day-ahead markets and real-time markets and operations. These issues include improvements in the opportunities for market participants to buy and sell energy across the seams with neighboring ISOs, greater opportunities for demand resources to participate in the New York markets, improved unit commitment, increases in the frequency with which market participants can make and adjust schedules and bids, improved modeling of constraints, and greater consistency among computer models and therefore of market prices. These changes will enhance market efficiency and system operations.

If the system works as expected, New York suppliers and consumers will face efficient prices for electricity, congestion, and ancillary services, prices that encourage the appropriate amounts and locations for production, consumption, and investment. As these real-time markets evolve, reliability standards may become less important. That is, the spot prices may provide sufficient incentives to reduce (or even eliminate) requirements for contingency reserves. Similarly, spot prices may provide the correct signals on the long-term need for new generation and demand-response programs (what FERC calls long-term resource adequacy).

REFERENCES

J. D. Chandley 2001, "A Standard Market Design for Regional Transmission Organizations," *The Electricity Journal* **14**(10), 27–53, December.

M. S. DePillis 2002, "New York ISO Real-Time Scheduling System and Some ISO New England Standard Market Design 1.0 Comparisons," NEPOOL Markets Committee Summer Meeting, ISO New England, Holyoke, MA, August 13.

A. Hartshorn, B. Kranz, and R. de Mello 2002, *Real-Time Commitment (RTC) and Real-Time Dispatch (RTD), Preliminary Design Considerations*, #A536, draft, New York Independent System Operator, Schenectady, NY, May 31.

E. Hirst 2001, *Real-Time Balancing Operations and Markets: Key to Competitive Wholesale Electricity Markets*, Edison Electric Institute, Washington, DC, and Project for Sustainable FERC Energy Policy, Alexandria, VA, April.

New York Independent System Operator 1999a, *NYISO Transmission and Dispatching Operations Manual*, Schenectady, NY, September 1.

New York Independent System Operator 1999b, *NYISO Day Ahead Scheduling Manual*, Schenectady, NY, September 2.

New York Independent System Operator 2002a, *Technical Bulletin #49, Multi-Pass Methodology of Security Constrained Unit Commitment*, Schenectady, NY, January 28.

New York Independent System Operator 2002b, *Real-Time Scheduling Architecture Concept of Operation*, draft, Schenectady, NY, May 31.

New York State Electric & Gas Corp. 2000, *Complaint of New York State Electric & Gas Corporation to Suspend Market Based Rates for Energy Markets and Request for Emergency Technical Conference*, Docket No. EL00-70-000, submitted to the Federal Energy Regulatory Commission, Binghamton, NY, April 24.

North American Electric Reliability Council 2002, *Reliability Assessment 2002–2011*, Princeton, NJ, October.

D. B. Patton 2002, *Summer 2002 Review of the New York Electricity Markets*, Potomac Economics, Ltd., Washington, DC, October 15.

D. B. Patton and M. T. Wander 2002, *2001 Annual Report on the New York Electricity Markets*, Potomac Economics, Ltd., Washington, DC, June.

S. Stoft 2002, *Power System Economics: Designing Markets for Electricity*, IEEE Press, Wiley-Interscience, New York, NY.

U.S. Federal Energy Regulatory Commission 2000, *New York Independent System Operator, Inc., Order on Tariff Filing and Complaint*, Docket Nos. ER00-3038-000 et al., Washington, DC, July 26.

U.S. Federal Energy Regulatory Commission 2002, *Notice of Proposed Rulemaking: Remedying Undue Discrimination through Open Access Transmission Service and Standard Electricity Market Design*, Docket No. RM01-12-000, Washington, DC, July 31.