ENERGY PRIMER
A Handbook of Energy Market Basics

Electric Power

Natural Gas

Transmission

Market Trading

LNG

A staff report of The Division of Energy Market Oversight
Office of Enforcement | Federal Energy Regulatory Commission
While the industry had historically traded electricity through bilateral transactions and power pool agreements, Order No. 888 promoted the concept of independent system operators (ISO). Along with facilitating open-access to transmission, an ISO would operate the transmission system independently of, and foster competition for electricity generation among, wholesale market participants. Several groups of transmission owners formed ISOs, some from existing power pools.

Close on the heels of Order No. 888, the Commission, in Order No. 2000, encouraged utilities to join regional transmission organizations (RTO) which, like an ISO, would operate the transmission systems and develop innovative procedures to manage transmission equitably. The Commission’s proceedings in Orders Nos. 888 and 2000, along with the efforts of the states and the industry, led to the voluntary organization of ISOs and RTOs. Each of the ISOs and RTOs subsequently developed a full scale energy and ancillary service market in which buyers and sellers could bid for or offer generation. The ISOs and RTOs used the bid-based markets to determine economic dispatch. Throughout the subsequent sections of the primer, the term RTO is used to stand for both ISOs and RTOs.

Major parts of the country operate under more traditional market structures, notably the West (excluding California) and the Southeast. Notably, two-thirds of the nation’s electricity load is served in RTO regions.
Electricity Demand

Americans use electricity for heat and light, to run machinery and to power a growing number of products such as televisions, radios, computers, hair dryers, and cell phones. This use has been increasing, reaching over 3,860 gigawatt-hours (GWh) of electricity in 2014. Demand dropped in 2009 with the recession, but has since recovered. Subsequent to the recession, the trend in use has been flatter.

The bulk of the electricity generated is sold to consumers, known as end-users or retail customers. Some consumers generate some or all of the power they consume. Some of the electricity sold to retail consumers is generated by integrated investor-owned utilities, federal entities, municipally owned and co-operatively owned utilities that sell the power directly to consumers. The rest of the electricity ultimately consumed by retail customers is bought and sold through wholesale electricity markets.

Demand Characteristics

Demand is often characterized as baseload or peak. Baseload is demand that occurs throughout the day or throughout the year. Refrigerators, for example, may create baseload demand. Peak load is demand that shows up during part of the day or year, all at the same time – heating or air conditioning, for example.

The amount of electricity consumed (demand) is continuously varying and follows cycles throughout the day and year. Regionally, electric demand may peak in either the summer or the winter. Spring and fall are typically shoulder months, with lower peak demand. Seasonal peaks vary regionally, although the highest levels of power load in almost all regions of the United States occur during heat waves, which are most acute during the daily peak load hours that occur during the late afternoon. However, a minority of regions reach their peak load when the weather is extremely cold. These are primarily areas with significant space-heating requirements and little summer air conditioning load. A majority of these systems are in the far northern areas of the United States, where air conditioning load is not significant. South Florida’s seasonal peak also occurs during the winter, when the population and tourism surges.

Daily demand typically peaks in the late afternoon, as commercial and domestic activities peak, and, in the winter, when lighting needs grow.

Electricity use also varies between weekdays and weekends. Commercial and industrial activities are lower on weekends and peoples’ noncommercial activities change with their personal schedules. The load on different weekdays also can behave differently. For example, Mondays and Fridays, being adjacent to weekends, may have different loads than Tuesday through Thursday. This is particularly true in the summer.

Because demand historically has not varied with price and because storage options are limited, generation must rise and fall to provide exactly the amount of electricity customers need. The cost of providing power typically rises as demand grows, and falls as demand declines, because higher levels of demand require activation of increasingly more expensive sources of power generation, and reductions as demand declines. As a result, power prices are typically highest during periods of peak demand. This causes system planners, power marketers and traders to all carefully track weather trends, economic growth and other factors to forecast power demand.

Demand Drivers

In general, the amount of electricity demanded is relatively insensitive to the price of electricity in the short-term (inelastic). One reason for this is that many customers – especially smaller customers – do not get price signals to which they can respond. Most residential customers are billed monthly on a preset rate structure. Large industrial customers, on the other hand, may receive real-time price signals.

Further, electricity is a necessity to most people and businesses. While they may be able to reduce their demand in the short-term – by turning down the thermostat or turning off
lights, for example – electricity consumers find it difficult to do without electricity altogether. There is little storage for electricity now and few realistic substitutes. Consequently, demand tends to drive price, especially when the system is stressed.

In the longer-term, options for reducing electricity use include switching to natural gas, installing insulation and implementing other energy efficiency measures. Larger consumers may consider building their own generation facilities.

Governments and businesses are also developing demand-response programs, which may provide reduced rates or other compensation to customers who agree to reduce load in periods of electric system stress.

Factors driving demand include demographics, climate and weather, economic activity and policies and regulations.

**Demographics**

Population levels affect demand, with greater population levels tending to increase electricity consumption. Shifts in population also affect regional demand. Population flight in the 1980s from northern industrial regions – the Rust Belt – to warmer climates in the South affected residential consumption patterns. In the 1990s, consumption in the South surpassed that in the Midwest, making it the region with the greatest electricity use.

**Climate and Weather**

Weather is one of the primary factors driving demand. General climatic trends drive consumption patterns and therefore the infrastructure needed to ensure reliable service. Cold weather and short days drive winter demand in northern regions. Southern regions rely more on electric space heating, and, thus, see demand rise in the winter, although demand typically peaks in the summer with air conditioning load. In the winter, lighting contributes to the occurrence of peaks during the seasonally dark early morning and early evening hours.

Weather also can have extreme short-term effects on electricity usage. A sudden cold snap can drive heating use up quickly and a heat wave can push air conditioning loads. Other, less obvious weather patterns affect demand – rain and wind, for example, may result in sudden cooling, affecting heating or air conditioning.

**Economic Activity**

The pattern of socioeconomic life affects the cycle of electricity use, with weekends and holidays showing a different pattern than weekdays. Demand typically rises as people wake up and go to work, peaking in the afternoon.

The overall level of economic activity also affects power demand. During periods of robust activity, loads increase. Similarly, loads drop during recessions. These changes are most evident in the industrial sector, where business and plants may close, downsize or eliminate factory shifts. In addition to reducing overall demand, these changes may affect the pattern of demand; for example, a factory may eliminate a night shift, cutting baseload use but continuing its use during peak hours. In some cases these effects can be significant.

**Quick Facts: Heating and Cooling Degree Days**

In the United States, engineers developed the concept of heating and cooling degree days to measure the effect of temperature on demand. Average daily temperatures are compared to a 65°F standard - those in excess of 65° yield cooling degree days; those below 65° yield heating degree days. A day with an average temperature of 66° would yield one cooling degree day.
Energy Policies and Regulations

State regulatory agencies set prices and policies affecting retail customer service. Some states are considering changes that would enable customers to receive more accurate price signals. They include, among other things, changing rate structures so that the rate varies with the time of day, or is even linked to the cost of providing electricity.

Efforts to reduce overall demand by improving energy efficiency are underway through several governmental and utility venues.

Retail Customer Mix

Most electric utilities serve different types of customers: residential, commercial and industrial. Each class uses electricity differently, resulting in a differing load profile, or the amount that each customer class uses and the daily shape of the load. If a consumer uses electricity consistently throughout the day and seasons, the load shape is flat, and the load will be baseload. Another consumer may use more at some times than others, resulting in baseload and peaks. Greater variability in demand is typically more expensive to serve, especially if the peak occurs at the same time other customers’ use peaks. Consequently, the mix of customer types affects a region’s overall demand.

Residential consumers form the largest customer segment in the United States at approximately 37 percent of electricity demand. Residential consumers use electricity for air conditioning, refrigerators, space and water heating, lighting, washers and dryers, computers, televisions and other appliances. Prices for residential service are typically highest, reflecting both their variable load shape and their service from lower-voltage distribution facilities, meaning that more power lines are needed to provide service to them.

Commercial use is the next largest customer segment at approximately 36 percent, and includes office buildings, hotels and motels, restaurants, street lighting, retail stores and wholesale businesses and medical, religious, educational and social facilities.

More than half of commercial consumers’ electricity use is for heating and lighting.

Industrial consumers use about 27 percent of the nation’s electricity. This sector includes, for example, manufacturing, construction, mining, agriculture and forestry operations. Industrial customers often see the lowest rates, reflecting their relatively flat load structure and their ability to take service at higher voltage levels.

Transportation demand for electricity stems primarily from trains and urban transportation systems. This is less than 1 percent of electricity demand.

Load Forecasting

Demand is constantly changing, challenging grid operators and suppliers responsible for ensuring that supply will meet demand. Consequently, they expend considerable resources to forecast demand. Missed forecasts, where actual demand differs significantly from the forecast, can cause wholesale prices to be higher than they otherwise might have been.

Forecasts are necessary as well for the variety of actions that must occur if sufficient supply is to be available in the immediate or long term: planning the long-term infrastructure needs of the system, purchasing fuel and other supplies and staffing, for example. Load forecasts are also extremely important for suppliers, financial institutions and other participants in electric energy generation, transmission, distribution and trading.

Load forecasting uses mathematical models to predict demand across a region, such as a utility service territory or RTO footprint. Forecasts can be divided into three categories: short-term forecasts, which range from one hour to one week ahead; medium forecasts, usually a week to a year ahead; and long-term forecasts, which are longer than a year. It is possible to predict the next-day load with an accuracy of approximately 1 to 3 percent of what actually happens. The accuracy of these forecasts is limited by the accuracy of the weather forecasts used in their preparation and the uncertainties of
human behavior. Similarly, it is impossible to predict the next year peak load with the similar accuracy because accurate long-term weather forecasts are not available.

The forecasts for different time horizons are important for different operations within a utility company. Short-term load forecasting can help to estimate transmission system power flows and to make decisions that can prevent overloading of transmission systems. Timely implementation of such decisions leads to the improvement of network reliability and to the reduced occurrences of equipment failures and blackouts. Forecasted weather parameters are the most important factors in short-term load forecasts; temperature and humidity are the most commonly used load predictors.

The medium- and long-term forecasts, while not precise, take into account historical load and weather data, the number of customers in different customer classes, appliances used in the area and their characteristics, economic and demographic data, and other factors. For the next-year peak forecast, it is possible to provide an estimated load based on historical weather observations. Long-term forecasts are used for system infrastructure planning and are meant to ensure that there are sufficient resources available to meet the needs of the expected future peak demand. These forecasts are made for periods extending 10 to 20 years into the future.

**Demand Response**

Electricity demand is generally insensitive to price, meaning that demand does not typically fall when wholesale prices rise. This occurs for several reasons, including that most end use consumers of electricity are not exposed to real-time electricity prices. However, some utilities and grid operators are developing ways to stimulate a response from consumers through demand-response programs.

Demand response (DR) is the reduction in consumption of electricity by customers from their expected consumption in response to either reliability or price triggers where the customer forgoes power use for short periods, shifts some high energy use activities to other times, or uses onsite generation. The programs may use price signals or incentives to prompt customers to reduce their loads. The signals to respond to electric power system needs or high market prices may come from a utility or other load-serving entity, an RTO or an independent DR provider. These programs are administered by both retail and wholesale entities.

DR has the potential to lower systemwide power costs and assist in maintaining reliability. It can be used instead of running power plants or to relieve transmission congestion. There can also be environmental benefits because peaking units tend to be costly and dirty to run.

Demand response rewards consumers for reducing load during certain market conditions and at specific times. However, it is difficult to measure and quantify this reduction. Measuring and verifying the reduction requires development of consumers’ baseline usage, against which their actual use is measured to determine the reduction in the event they are called to lessen their load. An accurate measure of their typical usage is important to prevent (or detect) gaming by participants.

**Demand-Response Programs**

Programs generally fall into three categories: curtailing, shifting or on-site generation.
Curtailing, or forgoing, involves reducing power use (load) during times of high prices or threats to reliability without making up the use later. For example, residential customers might turn off lights or raise thermostats during hot weather. Commercial facilities may turn off office equipment, lower building lighting or change thermostat settings by a few degrees.

Shifting involves moving or rescheduling high energy-use activities in response to high prices or DR program events to off-peak periods – evenings, nights or weekends. Industrial customers might reschedule batch production processes to evening hours or the next day. Commercial establishments may delay high-energy operations. Residential customers may wait until evening or night to use high-energy consuming appliances, such as clothes dryers or dishwashers. In shifting, the lost amenity or service is made up at a subsequent time.

On-site generation is when some customers may respond by turning on an on-site or backup emergency generator to supply some or all of their electricity needs. Although customers may have little or no interruption to their electrical usage, their net load and requirements from the power system are reduced. The ability to use on-site generation is most common for institutional customers, such as hospitals, large schools or data centers.

DR programs can be further distinguished by whether they are controlled by the system operator (dispatchable) or the customer (nondispatchable). Dispatchable demand response refers to programs that reduce customer energy use, such as direct load control of residential appliances or directed reductions to industrial customers. Dispatchable DR is used for reliability or economic reasons. Nondispatchable demand response lets the retail customer decide whether and when to reduce consumption in response to the price of power. It includes time-sensitive pricing programs based on rates that charge higher prices during high-demand hours and lower prices at other times.

As a result of technology innovations and policy directions, new types and applications of DR are emerging that encompass the use of smart appliances that respond in near real-time to price or other signals. These models may allow customers to respond more easily as they require little customer monitoring or interaction.

**Demand Response in Retail Markets**

Many states require utilities to use energy efficiency, DR or renewable resources. Energy Efficiency Resource Standards (EERS) in more than half of the states require utilities to achieve electric energy savings; many of these standards include peak load reduction targets. These mandates provide incentives for utilities to reduce customers’ energy consumption, such as mechanisms that decouple profits from the amount of electricity sold, or performance bonuses for utilities that meet or exceed reduction targets.

Some states are implementing dynamic pricing, in which retail rates change frequently to better reflect system costs. Time-based rates depend on advanced meters at customer premises that can record usage. In time-of-use programs, customers are charged different prices at different times, with hours of peak demand costing more than off-peak hours.

In real-time pricing (RTP) programs, customers are charged prices reflecting the immediate cost of power. Industrial or very large commercial customers are often on RTP tariffs.

Critical peak pricing (CPP) uses real-time prices at times of extreme system peak, and is restricted to a small number of hours annually but features prices higher than time-of-use prices during the critical peak. Consumers do not know in advance when a critical peak might be called. A CPP program for residential customers uses a carrot without the stick: critical-peak rebates. Participating customers get rebates on their bills for responding to utility price-signals, but are not penalized if they do not lower use in those hours.

**Wholesale Market Programs**

Retail programs may aid RTOs, although the RTO may not be able to invoke them or even see specifically the amount of response that occurs. Wholesale-level DR occurs in the
RTOs, which differ in how demand-response resources (DRR) may participate in their markets. Some RTOs permit DRR to participate in their markets as voluntary reliability resources. DRR also can participate in wholesale electricity markets as capacity resources and receive advance reservation payments in return for their commitment to participate when called. Resources that fail to perform when called are penalized. Additionally, some DRR bids into RTO day-ahead (DA) markets as energy resources, specifying the hours, number of megawatts and price at which they are willing to curtail. RTOs set minimum bid values.

Some of the RTO DR comes from individual entities; the rest is accumulated through third-party aggregators, or curtailment service providers (CSPs), who recruit customers too small to participate on their own, such as schools, commercial chains or groups of residential customers. In aggregating small customers, CSPs have increased customer participation in many wholesale reliability and emergency programs.

**Demand-Response Use in Planning and Operations**

Different DR programs can be used at various times to support planning and operations (see graphic). Energy efficiency programs that reduce baseload or peak demand over the long-term are incorporated into system planning. Dispatchable programs that are quickly implemented and targeted for short-term peak reductions – such as direct load control – lie on the other end of the spectrum, and are used in the moment of operation.
Electricity Supply and Delivery

Unlike many other products, electricity cannot be stored in any appreciable quantities. Further, electricity is a necessity for most consumers, whose use responds little to price changes. Finally, electric equipment and appliances are tuned to a very specific standard of power, measured as voltage. Deviations in voltage can cause devices to operate poorly or may even damage them. Consequently, the supply side of the electricity market must provide and deliver exactly the amount of power customers want at all times, at all locations. This requires constant monitoring of the grid and close coordination among industry participants.

Electricity service relies on a complex system of infrastructure that falls into two general categories: generation and the delivery services of transmission and distribution. Together, the power generation and high-voltage transmission lines that deliver power to distribution facilities constitute the bulk power system. Transmission and distribution facilities are also referred to as the power grid. These are coordinated and at times operated by a grid coordinator.

Electricity Supply and Delivery

Nationally, the grid is split into three main sections—the Western, Eastern and Texas Interconnections. These sections operate independently and have limited interconnections between them.

The nation, along with Canada and a small part of Mexico, is also divided into regional entities. The regional reliability entities fall under the purview of NERC, which was designated by FERC as the nation’s energy reliability organization and which develops standards, among other things, to ensure the grid’s reliability. The standards, once approved by FERC, must be met by all industry participants—the standards are mandatory and enforceable. Consequently, the grid is designed and operated to meet these standards.

NERC’s regions include:
- Florida Reliability Coordinating Council (FRCC)
- Midwest Reliability Organization (MRO)
- Northeast Power Coordinating Council (NPCC)
- Reliability First Corporation (RFC)
- SERC Reliability Corp. (SERC)
- Southwest Power Pool (SPP)
- Texas Reliability Entity (TRE)
- Western Electricity Coordinating Council (WECC)

NERC Regions

Source: North American Electric Reliability Corporation

Generation

Power generators are typically categorized by the fuel they use and subcategorized by their specific operating technology. The United States has more than 1,000 gigawatts (GW) of total generating capacity. Coal, natural gas and nuclear dominate the power generation market.

Power plants each have differing costs and operational characteristics, both of which determine when, where and how
plants will be built and operated.

Plant costs fall into two general categories: capital investment costs, which are amounts spent to build the plant, and operational costs, the amounts spent to maintain and run the plant. In general, there is a trade-off between these expenses: the most capital intensive plants are the cheapest to run – they have the lowest variable costs – and, conversely, the least capital intensive are more expensive to run – they have the highest variable cost. For example, nuclear plants produce vast amounts of power at low variable costs, but are quite expensive to build. Natural gas-fired combustion turbines are far less expensive to build, but are more expensive to run.

Grid operators dispatch plants – or, call them into service – with the simultaneous goals of providing reliable power at the lowest reasonable cost. Because various generation technologies have differing variable costs, plants are dispatched only when they are part of the most economic combination of plants needed to supply the customers on the grid. For plants operating in RTOs, this cost is determined by the price that generators offer. In other areas, it is determined by the marginal cost of the available generating plants.

Construction of different generating technologies is subject to a number of issues, including community concerns, regional emission restrictions and the availability of fuels or other necessary resources:

- Wind plants are generally built in areas with the appropriate meteorological conditions. In most cases, these sites are located in rural areas with limited transmission access. For example, in West Texas, the transmission lines connecting wind farms with consumer centers in Dallas and Houston can become overloaded, requiring generators to curtail production.
- Coal plants have environmental characteristics that limit both their siting and operations. Specifically, they emit NOx, SOx, particulates, mercury and substantially higher levels of CO2 than gas-fired plants. This has made financing these plants and siting them near urban centers difficult.
- There have been virtually no new nuclear plants built in the United States in the past 30 years. The technology of older plant designs became a source of concern following the accident at the Three Mile Island plant in the United States in 1977, the Chernobyl plant meltdown in Ukraine in 1986 and the Japanese earthquake, tsunami and nuclear plant destruction in 2011. New plant designs have been put forward and a few are under construction. The disposition of high-level radioactive waste remains an unresolved problem, and the waste remains at plant locations.

**Conventional Generation**

Natural gas power plants feature three major technologies, each with its distinct set of market advantages and limitations. They are steam boilers, gas turbines and combined cycle generators. Natural gas fuels nearly a third of electricity generation.

Steam boiler technology is an older design that burns gas in a large boiler furnace to generate steam at both high pressure and a high temperature. The steam is then run through a turbine that is attached to a generator, which spins and produces electricity. Typical plant size ranges from 300 MW to 1,000 MW. Because of their size and the limited flexibility that is inherent in the centralized boiler design, these plants require fairly long start-up times to become operational and are limited in their flexibility to produce power output beyond a certain range. Furthermore, these plants are generally not as economical or easy to site as some newer technologies – which explains why few have been built in recent years.

Gas turbines (GT) are small, quick-start units similar to an aircraft jet engine. These plants are also called simple cycle turbines or combustion turbines (CT). GTs are relatively inexpensive to build, but are expensive to operate because they are relatively inefficient, providing low power output for the amount of gas burned, and have high maintenance costs. They are not designed to run on a continuous basis and are used to serve the highest demand during peak periods, such as hot summer afternoons. GTs also run when there are system-wide shortages, such as when a power line or generator trips offline. GTs typically have a short operational life due to the wear-and-tear caused by cycling. The typical capacity of a
GT is 10-50 MW and they are usually installed in banks of multiple units.

**Combined cycle power plants (CCPPs)** are a hybrid of the GT and steam boiler technologies. Specifically, this design incorporates a gas-combustion turbine unit along with an associated generator, and a heat recovery steam generator along with its own steam turbine. The result is a highly efficient power plant. They produce negligible amounts of SO₂ and particulate emissions and their NOₓ and CO₂ emissions are significantly lower than a conventional coal plant. CCPPs, on average, require 80 percent less land than a coal-fired plant, typically 100 acres for a CCPP versus 500 acres for comparable coal plant, and CCPPs also use modest amounts of water, compared to other technologies.

**Coal plants** generate more than one-third of the electricity in the United States. These facilities tend to be large, baseload units that run continuously. They have high initial capital costs and are also somewhat complex in their design and operations. However, coal plants have low marginal costs and can produce substantial amounts of power. Most of the coal-fired plants in the United States are located in the Southeast and Midwest.

**Oil-fired plants** generally produce only a small amount of the total electricity generated in the U.S. power markets. These facilities are expensive to run and also emit more pollutants than gas plants. These plants are frequently uneconomic and typically run at low capacity factors. Like gas-fired generators, there are several types of units that burn oil; primarily, these are steam boilers and combustion turbines.

Generally, two types of oil are used for power generation: number 2 and number 6 (bunker) fuel oil. Number 2 is a lighter and cleaner fuel. It is more expensive, but because it produces fewer pollutants when burned, it is better for locations with stringent environmental regulations such as major metropolitan areas. Conversely, number 6 fuel oil is cheaper, but considered dirty because of its higher emissions. It is highly viscous (thick and heavy) and it comes from the bottom of the barrel in the refining process.

**Nuclear plants** provide roughly 20 percent of the nation’s electricity; there are close to 100 operating units with a total capacity of approximately 100 GW. These plants are used as baseload units, meaning that they run continuously and are not especially flexible in raising or lowering their power output. Nuclear plants have high capital and fixed costs, but low variable costs, which includes fuel cost. They typically run at full power for 18 or 24 months, which is the duration of a unit’s fuel cycle. At that point, they are taken off-line for refueling and maintenance. Outages typically last from 20 days to significantly longer, depending on the work needed.

**Renewable Generation**

Renewable resources use fuels that are not reduced or used up in the process of making electricity. They generally include biomass, geothermal, hydropower, solar, onshore and offshore wind, hydrokinetic projects, fuel cells using renewables and biogas.

Renewable generation, an important part of total U.S. capacity and generation, accounted for 13 percent of 2014 electricity generation. As total generation from all fuels has remained relatively constant in the recent years, renewable generation’s share has risen, spurred by state regulations and federal tax credits. As renewable generation becomes a larger per-
percentage of generation resources, integrating them into the operating power grid has presented challenges.

Wind and solar capacity have grown faster than other renewable resources in recent years. Wind capacity grew ten-fold (from approximately 6 GW to 65 GW) between 2003 and 2014. Utility-scale solar capacity more than tripled (from approximately 3.1 GW to 10 GW) between 2012 and 2014.

Additions are usually reported in megawatts of nameplate capacity. Actual capability varies from the nameplate for any unit type due to age, wear, maintenance or ambient conditions. But as renewable resources are often weather-dependent, their capacity factors – the ratio of average generation to the nameplate capacity for a specific period – have been lower (for example, approximately 30 percent), depending on the technology type, than for fossil-fuel-fired generation. Markets care about the difference between nameplate and capacity factor values when they evaluate capacity available to cover expected load. Capacity factors have risen with technological innovation and improved manufacturing processes.

**Wind generation** is among the fastest-growing renewable resources, in part due to cost declines and technology improvements as well as earlier receipt of federal tax credits. Increases in average hub heights and rotor diameters have increased average wind turbine capacity.

Because the best wind resources are often far from load centers, obtaining sufficient transmission presents a challenge to delivering its output. Other market challenges for future wind development include its variable output, which is often inversely correlated to demand (seasonally and daily); system operators’ inability to dispatch wind resources to meet load increases; difficulties related to accurately forecasting its ramping; and the need for companion generation (usually fossil-fueled) to be available to balance wind generation when the wind is not blowing.

**Geothermal generation** taps into reservoirs of steam and hot water deep beneath the earth’s surface to produce power. The best resources are in the intermountain West. Geothermal potential is determined by thermal conductivity, thickness of sedimentary rock, geothermal gradient, heat flow and surface temperature. Geothermal generation increased 14 percent from 2006 to 2014, but it decreased from 15 to 6 percent of non-hydro renewable output, due to the growth of other renewables. California hosts more than 80 percent of U.S. operating capacity.

**Solar generation** transforms sunlight into electricity using one of two technologies: photovoltaic (PV) or concentrating solar power (CSP). PV modules, or panels, transform sunlight directly into power using silicon wafers or nonsilicon thin-film technologies. They can be installed on roofs of buildings or at ground-level PV farms. CSP plants use a two-step process to transform the sun’s energy. First, mirrors direct sunlight towards a receiver that captures the heat. CSP then employs a thermal process to create steam, driving an engine or turbine to produce electricity. CSP plants, which are dispatchable, can include low-cost energy storage that extends their availability later in peak hours.

PV growth has increased greatly as a result of policy incentives and cost declines. Annual PV installations increased nearly
tenfold from 2009 to 2013 as PV system costs decreased. PV growth has been relatively concentrated; 10 states had 90 percent of PV capacity in 2014, while California alone had over half.

By the end of 2014, 1,760 MW of CSP was operational. Seven western and southwestern states have extensive CSP potential: Utah, New Mexico, Arizona, Nevada, Texas, California and Colorado. Developing that potential will require overcoming challenges of siting, transmission and the need for extensive water supplies to clean mirrors.

**Hydroelectric generation** is powered by the kinetic energy of falling water that drives turbine generators, which convert the energy into electricity. There are two types of hydroelectric projects: conventional and pumped storage. Conventional projects, which use a dam in a waterway, can operate in a run-of-river mode, in which water outflow from the project approximates inflow, or in a peaking mode, in which the reservoir is mostly drained to generate power during peak periods when energy is more valuable. Pumped storage projects use bodies of water at two different elevations. Water is pumped into elevated storage reservoirs during off-peak periods when pumping energy is cheaper; the water is then used to generate power during peak periods as it flows back to the lower elevation reservoir. Pumped storage is the only significant commercially deployed electricity storage technology available today.

**Biomass generation** includes production from many waste by-products, such as agricultural residues, landfill gas, municipal solid waste and wood resources. The largest biomass category is wood waste, burned for heat and power in the lumber, pulp and paper industries. Challenges to biomass production include impacts on food supplies (for example, converting corn into ethanol), conserving natural resources and minimizing water pollution. State policies on renewable generation differ on eligibility of biomass technologies.

**Biogas energy** is created through the anaerobic (without oxygen) bacterial decomposition of manure, which is turned into a gas containing 60-70 percent methane. Biogas recovery can be installed at farms anywhere, used to run farm operations and reduce methane emissions from natural manure decomposition.

---

**Renewable Energy Policies**

Renewable development is frequently tied to policies promoting their use because of their higher cost relative to other technologies. Financial incentives include tax credits, low-cost loans, rebates or production incentives. Federal funding of research and development (R&D) has played an important role in lowering costs or reducing the time it takes for renewable technologies to become commercially viable.

Congress has provided tax incentives to spur renewable resource investments. Originally enacted in 1992, wind, biomass, geothermal, and other forms of renewable generation have been able to receive federal production tax credits (PTC) based on a facility’s production. An inflation-adjusted credit, the PTC generally has a duration of 10 years from the date the facility goes online. The credit was initially set at 1.5¢/kilowatt hour (kWh) and its value in 2013 was 2.3¢/kWh or 1.1¢/kWh, depending on the type of qualifying resource. The PTC has been renewed and expanded numerous times, most recently at the end of 2014.

Another form of tax credit for renewables, including solar and other select renewable energy projects, has been a federal investment tax credit (ITC). The ITC has generally been for 30 percent of a project’s equipment and construction costs.

Following the financial crisis of 2008, the American Reinvestment and Recovery Act (ARRA) provided developers with another option for projects that began construction by the end of 2010 – they could apply for Treasury-administered cash grants, which monetized the ITC’s value up front. ARRA funds helped support renewable energy research and development and aided capacity growth in 2009, despite the economic downturn.

State renewable portfolio standards (RPS) and renewable energy standards (RES) have been significant drivers in the
growth of investment in renewable generation. An RPS requires a certain percentage of energy sales (MWh) to come from renewable resources. Percentages usually increase incrementally from a base year to an ultimate target. Currently, 29 states plus Washington, D.C., have an RPS and eight states have renewable goals without financial penalties for nonachievement. As utilities build more renewable-powered generation, the markets in which they participate continue to address the integration of renewable output into their day-ahead and real-time operations and model expected growth as part of their long-term transmission-planning processes.

To encourage the development of distributed generation (DG), or the production of electricity at the site of consumption, and solar power, 16 states plus Washington, D.C., created RPS carve-outs or set-asides to give an extra boost to these resources, which are not yet cost-competitive with other renewables.

Renewable energy certificates (RECs) allow state regulators to track compliance with mandatory RPS targets or verify progress in voluntary state renewable programs. They also allow compliance entities to purchase credits – subject to state imposed limits on amount and price - if they have not generated or bought enough renewable energy to meet their annual requirements. Each reported megawatt-hour (MWh) of eligible generation results in a system-issued REC with a unique identification number to prevent double-counting. Each REC includes attributes such as generator location, capacity, fuel-type and source, owner and the date when operations began.

States and local utilities offer a variety of financial incentives for renewable energy to complement policy mandates. These include tax credits for in-state manufacture of renewable energy equipment, consumer rebates for purchase and installation of renewable generation or production incentives. Production incentives include extra credits for solar output based on RPS solar set-asides and feed-in tariffs.

Seven states mandate feed-in-tariffs (FITs) to support their energy and environmental goals. Also called feed-in rates or advanced renewable incentives, these programs typically are designed to encourage development of new small- and medium-sized renewable generation projects by residential and independent commercial developers.

FITs require utilities to buy the renewable generation at a fixed rate that is higher than that provided to other generators, under multiyear contracts. This enables smaller distributed renewable generators to avoid having to participate in renewable portfolio standard (RPS) auctions or other competitive procurements and compensates them for more expensive technologies. The utility passes the costs of the program to its customers.

Transmission

The alternating current (AC) power grid operates like an interconnected web, where, with a few exceptions, the flow of power is not specifically controlled by the operators on a line-by-line basis. Instead, power flows from sources of generation to consumers across any number of lines simultaneously, following the path of least resistance. There are a limited number of direct current (DC) lines, which are set up as specific paths with definite beginning and end points for scheduling and moving power. These lines are controllable by operators
and have other characteristics that make them attractive to grid planners and operators, such as providing greater grid stability and lower line losses. However, DC lines cost significantly more than AC lines to construct. Consequently, DC lines are typically built for certain specialized applications involving moving large amounts of power over long distances, such as the Pacific Intertie, which extends between the Northwest and California.

Transmission lines provide a certain amount of resistance to the flow of power as electricity travels through them. This resistance is not unlike the wind resistance that a car must overcome as it travels along a highway. The resistance in power lines creates losses: the amount of power injected into a power line diminishes as it travels through the line. The amount of these losses is contingent on many factors, but typically equals several percent of the amount put into the system.

**Transmission Service**

FERC requires that public utilities that own transmission lines used in interstate commerce offer transmission service on a nondiscriminatory basis to all eligible customers. The rates and terms of service are published in each utility’s Open Access Transmission Tariff (OATT). One type of service is point-to-point service. This service involves paying for and reserving a fixed quantity of transmission capacity and moving power up to the reservation amount from one location, the point of receipt (POR), to another location, the point of delivery (POD). Depending on availability, customers may purchase point-to-point service for durations of one hour to multiple years. The price for the service is cost-based and published in the OATT. In cases where there are multiple parties desiring transmission, it is allocated to the party willing to purchase it for the longest period of time. Capacity reassignment is the term for the resale of point-to-point transmission capacity in the secondary market.

Transmission holders may want to sell capacity in the secondary market because it is unneeded, or to make a profit. Capacity reassignment has been permitted since 1996. Beginning in 2007, resellers have been permitted to charge market-based prices for capacity reassignments, as opposed to the original cost-based price at which they purchased the capacity. The number of capacity reassignments increased from around 350 in 2007 to 30,000 in 2012. Most of the transactions were hourly, although capacity can also be reassigned on a daily, monthly or yearly basis.

If the market price of energy is greater at the POD than at the POR, the transmission has value. The transmission holder can capture this value by using the transmission – buying energy at the POR, moving it to the POD and selling it. Alternatively, the transmission holder can sell the transmission through a capacity reassignment. Thus, the price of a capacity reassignment should be equal to the expected price differential between the POD and the POR.

**Grid Operations**

Grid operators dispatch their systems using the least costly generation consistent with the constraints of the transmission system and reliability requirements. The dispatch process occurs in two stages: day-ahead unit commitment, or planning for the next day’s dispatch, and economic dispatch, or dispatching the system in real time.

**Day-Ahead Unit Commitment**

In the unit commitment stage, operators decide which generating units should be committed to be online for each hour, typically for the next 24-hour period. This is done in advance of real-time operations because some generating units require several hours lead time before they are brought online. In selecting the most economic generators to commit, operators take into account forecast load requirements and each unit’s physical operating characteristics, such as how quickly output can be changed, maximum and minimum output levels and the minimum time a generator must run once it is started. Operators must also take into account generating unit cost factors, such as fuel and nonfuel operating costs and the cost of environmental compliance.
Also, forecast conditions that can affect the transmission grid must be taken into account to ensure that the optimal dispatch can meet load reliably. This is the security aspect of commitment analysis. Factors that can affect grid capabilities include generation and transmission facility outages, line capacities as affected by loading levels and flow direction and weather conditions. If the security analysis indicates that the optimal economic dispatch cannot be carried out reliably, relatively expensive generators may have to replace less-expensive units.

**System and Unit Dispatch**

In the system dispatch stage, operators must decide in real time the level at which each available resource from the unit commitment stage should be operated, given the actual load and grid conditions, so that overall production costs are minimized. Actual conditions will vary from those forecast in the day-ahead commitment, and operators must adjust the dispatch accordingly. As part of real-time operations, demand, generation and interchange (imports and exports) must be kept in balance to maintain a system frequency of 60 hertz. This is typically done by automatic generation control (AGC) to change the generation dispatch as needed.

The chart below is a depiction of the market supply curve of the power plants for the New York Independent System Operator (NYISO). This is also commonly called the supply stack. In it, all of the plants in the New York market are shown sorted according to their marginal cost of production. Their cost of production is shown on the vertical axis. The cheapest ones to run are to the left and the most expensive to the right.

Dispatch in New York, for example, first calls on wind plants, followed successively by hydro, nuclear and coal-, gas- and oil-fired generators. This assumes that the plants have sufficient resources – enough wind for the wind powered generators or enough river flow for the hydroelectric plants, for example – and that sufficient transmission capability exists to deliver plant output and meet reliability needs.

**Market Supply Curve for NYISO (Illustrative)**
In addition, transmission flows must be monitored to ensure that flows stay within voltage and reliability limits. If transmission flows exceed accepted limits, the operator must take corrective action, which could involve curtailing schedules, changing the dispatch or shedding load. Operators may check conditions and issue adjusted dispatch instructions as often as every five minutes.

**Ancillary Services**

Ancillary services maintain electric reliability and support the transmission of electricity. These services are produced and consumed in real-time, or in the very near term. NERC and regional entities establish the minimum amount of each ancillary service that is required for maintaining grid reliability.

**Regulation** matches generation with very short-term changes in load by moving the output of selected resources up and down via an automatic control signal, typically every few seconds. The changes are designed to maintain system frequency at 60 hertz. Failure to maintain a 60-hertz frequency can result in collapse of an electric grid.

**Operating reserves** are needed to restore load and generation balance when a generating unit trips off line. Operating reserves are provided by generating units and demand resources that can act quickly, by increasing output or reducing demand, to make up a generation deficiency. There are three types:

- **Spinning reserves** are primary. To provide spinning reserve a generator must be on line (synchronized to the system frequency) with some unloaded (spare) capacity and be capable of increasing its electricity output within 10 minutes. During normal operation these reserves are provided by increasing output on electrically synchronized equipment or by reducing load on pumped storage hydroelectric facilities. Synchronized reserve can also be provided by demand-side resources.

- **Nonspinning reserves** come from generating units that can be brought online in 10 minutes. Nonspinning reserve can also be provided by demand-side resources.

- **Supplemental reserves** come from generating units that can be made available in 30 minutes and are not necessarily synchronized with the system frequency. Supplemental reserves are usually scheduled in the day-ahead market, allowing generators to offer their reserve energy at a price, thus compensating cleared supply at a single market clearing price. This only applies to ISO/RTOs, and not all reliability regions have a supplemental reserve requirement.

**Black start** generating units have the ability to go from a shutdown condition to an operating condition and start delivering power without any outside assistance from the electric grid. Hydroelectric facilities and diesel generators have this capability. These are the first facilities to be started up in the event of a system collapse or blackout to restore the rest of the grid.

**Reactive power:** Electricity consists of current, the flow of electrons, and voltage, the force that pushes the current through the wire. Reactive power is the portion of power that establishes and maintains electric and magnetic fields in AC equipment. It is necessary for transporting AC power over transmission lines, and for operating magnetic equipment, including rotating machinery and transformers. It is consumed
by current as it flows. As the amount of electricity flowing on a line increases, so does the amount of reactive power needed to maintain voltage and move current. Power plants can produce both real and reactive power, and can be adjusted to change the output of both. Special equipment installed on the transmission grid is also capable of injecting reactive power to maintain voltage.

**Weather**

Weather is the single greatest driver of electricity demand and, thus, is a major factor in grid operations. System operators therefore rely heavily on weather forecasts to ensure they have the right generation in the right locations to run the grid reliably.

Weather affects grid operations in other ways, as well. Primary among these is on the productivity of certain types of power generators: wind and hydroelectric. Wind turbines’ power output changes with wind availability and speed, which affects cost of wholesale power.

Hydroelectric plants rely on rain and snowfall to provide the river flow needed for their output. Geographically, this is most important in the Pacific Northwest, where seasonal hydro plant output is a critical source of power. Rain and the melting of winter snowpack feed the Columbia and Snake river systems. Surplus power from these generators is typically exported to California to help meet summer peak demand and provide a combination of increased reliability and lower prices.

Temperature can also affect the output of other power plants and capacity of transmission lines. Specifically, thermal plants that use a turbine – coal, gas, oil and nuclear plants – become less efficient at higher temperatures. Additionally, the capacity of transmission lines is limited by heat because the conductive material used in fabrication becomes more electrically resistant as they heat up, limiting their throughput.

---

**Wholesale Electricity Markets and Trading**

**Overview**

Markets for delivering power to consumers in the United States are split into two systems: traditional regulated markets and market-regulated markets run by RTOs.

In general, RTOs use their markets to make operational decisions, such as generator dispatch. Traditional systems rely on management to make those decisions, usually based on the cost of using the various generation options.

Trading for power is also split into over-the-counter (OTC) or bilateral transactions, and RTO transactions. Bilateral transactions occur in both traditional systems and in RTO regions, but in different ways.

Pricing in both RTO and traditional regions incorporate both cost-of-service and market-based rates.

**Bilateral Transactions**

Bilateral or OTC transactions between two parties do not occur through an RTO. In bilateral transactions, buyers and sellers know the identity of the party with whom they are doing business.

Bilateral deals can occur through direct contact and negotiation, through a voice broker or through an electronic brokerage platform, such as the IntercontinentalExchange (ICE). The deals can range from standardized contract packages, such as those traded on ICE, to customized, complex contracts known as structured transactions.

Whether the trade is done on ICE, directly between parties or through another type of broker, the trading of standard physical and financial products, such as next-day on-peak firm or swaps, allows index providers to survey traders and publish price indexes. These indexes provide price transparency.
Physical bilateral trades involving the movement of the energy from one point to another require that the parties reserve transmission capacity to move the power over the transmission grid. Transmitting utilities are required to post the availability of transmission capacity and offer service on an Open Access Same-Time Information System (OASIS) website. Traders usually reserve transmission capacity on OASIS at the same time they arrange the power contract.

When it comes time to use the reservation to transfer power between balancing authorities, one of the parties to the transaction submits an eTag electronically to Open Access Technology International (OATI), NERC’s eTag contractor. OATI will process the tag and send it to all parties named on the eTag. This ensures the orderly transfer of energy and provides transmission system operators the information they need to institute curtailments as needed. Curtailments may be needed when a change in system conditions reduces the capability of the transmission system to move power and requires some transactions to be cut or reduced.

Bilateral physical transactions conducted in RTOs are settled in standardized contracts, and load is served through the power dispatched by the RTO. The RTO then settles bilateral transactions based on the prices in the contracts and the prices that occurred in the RTO markets.

**Cost-Based Rates**

Cost-based rates are used to price most transmission services and some electricity when the Commission determines that market-based rates are not appropriate, or when an entity does not seek market-based rate authority. Cost-based rates are set to recover costs associated with providing service and give a fair return on capital. Cost-based rates are typically listed in a published tariff.

The following are major inputs to setting cost-based electricity rates:

- Determining used-and-useful electricity plants. This may include generation facilities, transmission facilities, distribution plants and office and related administration facilities.
- Determining expenses from the production, transmission and distribution of electricity, including fuel and purchased power, taxes and administrative expenses.
- Establishing a fair return on capital, known as the cost of capital. This includes determining the cost of debt, common equity, preferred stock and commercial paper and other forms of short-term borrowing such as lines of credit used to finance projects and provide cash for day-to-day operations.
- Allocating electric plant and other expenses among various customer classes and setting the rate structure and rate levels.

**Market-Based Rates**

Under market-based rates, the terms of an electric transaction are negotiated by the sellers and buyers in bilateral markets or through RTO market operations. The Commission grants market-based rate authority to electricity sellers that demonstrate that they and their affiliates lack or have adequately mitigated horizontal market power (percent of generation owned relative to total generation available in a market), and vertical market power (the ability to erect barriers to entry or influence the cost of production for competitive electricity suppliers). Wholesale sellers who have market-based rate authority and who sell into day-ahead or real-time markets administered by a RTO do so subject to the specific RTO market rules approved by the Commission and applicable to all market participants. Thus, a seller in such markets not only must have an authorization based on analysis of that individual seller’s market power, but it must abide by additional rules contained in the RTO tariff.

**Supplying Load**

Suppliers serve customer load through a combination of self-supply, bilateral market purchases and spot purchases. In addition to serving load themselves, load-serving entities (LSEs) can contract with others to do so. The choices are:
Self-supply means that the supplying company generates power from plants it owns to meet demand. Supply from bilateral purchases means that the load-serving entity buys power from a supplier. Supply from spot RTO market purchases means the supplying company purchases power from the RTO.

LSEs' sources of energy vary considerably. In ISO-NE, NYISO and CAISO, the load-serving entities divested much or all of their generation. In these circumstances, LSEs supply their customers' requirements through bilateral and RTO market purchases. In PJM, MISO and SPP, load-serving entities may own significant amounts of generation either directly or through affiliates and therefore use self-supply as well as bilateral and RTO market purchases.

**Traditional Wholesale Electricity Markets**

Traditional wholesale electricity markets exist primarily in the Southeast, Southwest and Northwest where utilities are responsible for system operations and management, and, typically, for providing power to retail consumers. Utilities in these markets are frequently vertically integrated – they own the generation, transmission and distribution systems used to serve electricity consumers. They may also include federal systems, such as the Bonneville Power Administration, the Tennessee Valley Authority and the Western Area Power Administration. Wholesale physical power trading typically occurs through bilateral transactions. Utilities in traditional regions have the following responsibilities:

- Generating or obtaining the power needed to serve customers (this varies by state)
- Ensuring the reliability of its transmission grid
- Balancing supply and demand instantaneously
- Dispatching its system resources as economically as possible
- Coordinating system dispatch with neighboring balancing authorities
- Planning for transmission requirements within the utility’s footprint
- Coordinating its system development with neighboring systems

**Regional Electricity Markets**

Two-thirds of the population of the United States is served by electricity markets run by regional transmission organizations or independent system operators (this primer uses RTO to stand for both RTOs and ISOs). The main distinction between RTO markets and their predecessors (such as vertically integrated utilities, municipal utilities and co-ops) is that RTO markets deliver reliable electricity through competitive market mechanisms.

The basic functions of an RTO include the following:

- Ensure the reliability of the transmission grid
- Operate the grid in a defined geographic footprint
- Balance supply and demand instantaneously
- Operate competitive nondiscriminatory electricity markets
- Provide nondiscriminatory interconnection service to generators
- Plan for transmission expansion on a regional basis

In performing these functions, RTOs have operational control of the transmission system, are independent of their members, transparently manage transmission congestion, coordinate the maintenance of generation and transmission system,
and oversee a transmission planning process to identify needed upgrades in both the near- and long-term.

RTOs do not own transmission or generation assets perform the actual maintenance on generation or transmission equipment, or directly serve end use customers.

Currently, seven RTOs operate in the United States, listed below in order of the size of their peak load:

- PJM Interconnection (PJM), 165 GW (summer of 2011)
- Midcontinent ISO (MISO), 126 GW (summer of 2011)
- Electric Reliability Council of Texas (ERCOT), 68 GW (summer of 2011)
- California ISO (CAISO), 50 GW (summer of 2006)
- Southwest Power Pool (SPP), 48 GW (summer of 2011)
- New York ISO (NYISO), 34 GW (summer of 2013)
- New England ISO (ISO-NE), 28 GW (summer of 2006)

**RTO Markets and Features**

RTO market operations encompass multiple services that are needed to provide reliable and economically efficient electric service to customers. Each of these services has its own parameters and pricing. The RTOs use markets to determine the provider(s) and prices for many of these services. These markets include the day-ahead energy market (sometimes called a Day 2 market), real-time energy market (sometimes called a Day 1 or balancing market), capacity markets (designed to ensure enough generation is available to reliably meet peak power demands), ancillary services markets, financial transmission rights (contracts for hedging the cost of limited transmission capability) and virtual trading (financial instruments to create price convergence in the day-ahead and real-time markets).

**RTO Energy Markets**

All RTO electricity markets have day-ahead and real-time markets. The day-ahead market schedules electricity production and consumption before the operating day, whereas the real-time market (also called the balancing market) reconciles any differences between the schedule in the day-ahead market and the real-time load while observing reliability criteria, forced or unplanned outages and the electricity flow limits on transmission lines.

The day-ahead energy market produces financially binding schedules for the production and consumption of electricity one day before its production and use (the operating day). The purpose of the day-ahead market is to give generators and load-serving entities a means for scheduling their activities sufficiently prior to their operations, based on a forecast of their needs and consistent with their business strategies.

In day-ahead markets, the schedules for supply and usage of energy are compiled hours ahead of the beginning of the operating day. The RTO then runs a computerized market model that matches buyers and sellers throughout the geographic market footprint for each hour throughout the day. The model then evaluates the bids and offers of the participants, based on the power flows needed to move the electricity throughout the grid from generators to consumers. Additionally, the model must account for changing system capabilities that occur based on weather and equipment outages, plus rules and procedures that are used to ensure system reliability. The market rules dictate that generators submit supply offers and loads submit demand bids to the RTO by a deadline that is typically in the morning of the day-ahead scheduling. Typically, 95 percent of all energy transactions are scheduled in the day-ahead market, and the rest scheduled in real-time.

Generation and demand bids that are scheduled by the day-ahead market are settled at the day-ahead market prices. Inputs into setting a day-ahead market schedule include:

- Generator offers to sell electricity each hour
- Bids to buy electricity for each hour submitted by load-serving utilities
- Demand-response offers by customers to curtail usage of electricity
- Virtual demand bids and supply offers
- Operational information about the transmission grid and generating resources, including planned or known transmission and generator outage, the physical characteristics

Energy Primer | 59
of generating resources including minimum and maximum output levels and minimum run time and the status of interconnections to external markets.

The real-time market is used to balance the differences between the day-ahead forecast and the actual real-time load. The real-time market is run hourly and in 5-minute intervals and clears a much smaller volume of energy and ancillary services than the day-ahead market, typically accounting for only 5 percent of scheduled energy. For generators, the real-time market provides additional opportunities for offering energy into the market. Megawatts over- or under-produced relative to the day-ahead commitments are settled at real-time prices.

Real-time market prices are significantly more volatile than the day-ahead market prices. This stems from demand uncertainty, transmission and generator forced outages and other unforeseen events. Because the day-ahead market generally is not presented with these events, it produces more stable prices than in real-time. Also, because the volumes in the real-time market are much smaller, there is an increased likelihood of supply and demand imbalances, which lead to both positive and negative price movements.

RTOs use markets to deal with transmission constraints through locational marginal pricing (LMP). The RTO markets calculate a LMP at each location on the power grid. The LMP reflects the marginal cost of serving load at the specific location, given the set of generators that are dispatched and the limitations of the transmission system. LMP has three elements: an energy charge, a congestion charge and a charge for transmission system energy losses.

If there are no transmission constraints, or congestion, LMPs will not vary significantly across the RTO footprint. Transmission congestion occurs when there is not enough transmission capacity for all of the least-cost generators to be selected. The result is that some more expensive generation must be dispatched to meet demand, units that might not otherwise run if more transmission capacity were available.

When there are transmission constraints, the highest variable cost unit that must be dispatched to meet load within transmission-constrained boundaries will set the LMP in that area. All sellers receive the LMP for their location and all buyers pay the market clearing price for their location.

The primary means used for relieving transmission congestion constraints is by changing the output of generation at different locations on the grid. The market-based LMP sends price signals that reflect congestion costs to market participants. That is, LMPs take into account both the impact of specific generators on the constrained facility and the cost to change (redispacht) the generation output to serve load. This change...
in dispatch is known as security constrained redispatch.

This redispatch could be implemented by using nonmarket procedures such as transmission loading relief (TLR). NERC established the TLR process for dealing with reliability concerns when the transmission network becomes overloaded and power flows must be reduced to protect the network. A TLR is used to ration transmission capacity when the demand for transmission is greater than the available transmission capacity (ATC). The rationing is a priority system that cuts power flows based on size, contractual terms and scheduling.

- Establish a pricing structure for operating reserves that would raise prices as operating reserves grow short (demand curve)
- Set the market-clearing price during an emergency for all supply and demand response resources dispatched equal to the payment made to participants in an emergency demand-response program

Reliability must-run (RMR) units are generating plants that would otherwise retire but the RTO has determined they are needed to ensure reliability. They could also be units that have market power due to their location on the grid. RTOs enter into cost-based contracts with these generating units and allocate the cost of the contract to transmission customers. In return for payment, the RTO may call on the owner of an RMR generating unit to run the unit for grid reliability. The payment must be sufficient to pay for the cost of owning and maintaining the unit even if it does not operate.

Transmission upgrades can also reduce the need for RMR units by increasing generation deliverability throughout the RTO.

RTO Capacity Markets

RTOs, like other electric systems, are required to maintain adequate generation reserves to ensure that sufficient generation and demand-resource capacity are available to meet load and reliability requirements. LSEs have typically satisfied their reserve obligations with owned generation or bilateral contracts with other suppliers. Some RTOs have mechanisms to obtain capacity commitments, such as capacity auctions and capacity payments.

Most RTOs run a capacity market to allow LSEs a way to satisfy their reserve obligation. These markets cover short-term capacity, such as a month, season or year. PJM and ISO-NE run capacity auctions up to three years prior to when the capacity is needed. The near-term focus is consistent with providing payments to existing generation, or generation such as combustion turbines that can be sited and built within three years.