©ISTOCKPHOTO.COM/SIBERIANART

By Benjamin F. Hobbs
and Shmuel S. Oren

Three Waves of U.S. Reforms

Following the Path of Wholesale Electricity Market Restructuring

Digital Object Identifier 10.1109/MPE.2018.2873952
Date of publication: 7 January 2019

NEW GENERATION, STORAGE, DEMAND MANAGEMENT, AND SMART grid technologies have the potential to make the grid of 2050 very different from today's power system. Can market designers keep up, and what do they need to pay attention to now? Before offering a few answers to these questions, we first give some historical perspective on the creation of markets for electric power and how those markets have been adapted to new needs.

The power industry is now undergoing dramatic changes, following many transitions that we have already experienced. In the 1960s, nearly all power was provided by vertically integrated utilities. They confidently planned ever-larger coal and nuclear generators (including floating nuclear plants), whose prudently incurred costs would be guaranteed by ratepayers. At least one industry group concluded that its greatest environmental challenge was the ugliness of overhead distribution lines; climate change meant global cooling and was irrelevant to planners and regulators.

All of that has now changed beyond recognition in most (but not all) of the world. Future changes in the industry will differ in type but not in magnitude from those past changes. If market designs adhere to the principles of open access, technology neutrality, recognition of network realities, and financial rewards for

1540-7977/19©2019IEEE

providing what the system really needs when and where it needs it, then the power system will be well placed to adapt to whatever technology and social changes are coming.

The First Wave of Reforms: Unbundling and Debunking the Natural Monopoly Myth

The first introduction of competition into the electric power sector of the U.S. economy is often attributed to the 1978 Public Utilities Regulatory Policy Act (PURPA). That act required electric utilities, which were most commonly vertically integrated monopolies within their service territories, to buy power produced by small (lower than 50 MW) and renewable power sources.

PURPA came about at a time when Washington, D.C., was abuzz about deregulation. Economist Fred Kahn of Cornell, who was President Jimmy Carter's chief economic advisor, is largely credited with first promoting the liberalization of markets away from heavily regulated or publicly owned firms. The basic idea was that freer markets could spur companies to offer a wider range of products and lower costs. Policy makers deregulated and restructured railroads, telecommunications, air travel, shipping, securities markets, and gas and water utilities. In the 1990s, policy makers, consumers, traders, and academics sought to bring the same revolution to electric power.

However, PURPA and the subsequent reforms of the 1990s were not the first time that competition was introduced to power provision. Britain's Central Electricity Board created an auction in the 1930s in which power was contracted from the providers who could offer it most efficiently, contributing to a halving of heat rates in that decade. In the United States, despite Samuel Insull's success in promoting the regulatory compact between monopoly power providers and states that agreed to regulate them, distribution-level competition was not completely eliminated. In some cities, rival distribution companies ran lines in parallel to poach each other's customers; there is some evidence that lower rates resulted where there was such competition. Although post–World War II nationalization displaced Britain's competitive market and distribution-level competition eventually disappeared in the United States, these examples were an inspiration to U.S. reformers in the 1990s.

Prior to the 1990s, free-market ideology in the United Kingdom and Chile led to the first wholesale restructurings of the power sector, unburdened by any felt need for careful analysis of benefits and costs. In British Prime Minister Margaret Thatcher's view, there was simply "no other alternative," given, e.g., how the nationalized industry failed to keep the lights on during the coal strikes earlier in the 1970s. At the same time, in the United States, academics were establishing the intellectual underpinnings for restructuring. Fred Schweppe of the Massachusetts Institute of Technology (MIT) and his students mathematically described how prices could vary over time and from bus to bus based on the instantaneous marginal cost of delivering power and how such spot prices could incentivize building the right amount, type, and place of generation capacity and even transmission (Schweppe et al., 1988). A great irony is that Schweppe first proposed that this be used to motivate consumers to modify their behavior; only later was this idea extended to supply. The irony is that, in today's markets, the demand side of the market remains largely insulated from prices that reflect actual system conditions in time and space. Meanwhile, Schweppe's MIT colleagues Paul Joskow and Richard Schmalensee (1983) described how institutions could be defined to promote competition in distribution, transmission, and especially generation.

But it was not the earlier examples of competition or the intellectual frameworks proposed by Schweppe et al. that were chiefly responsible for spurring the U.S. restructuring efforts of the 1990s. Instead, the large and obvious decline in the cost of rail shipments, natural gas, telephone calls, and air travel was a major inspiration. Another driver was that large customers chafed under rising electric rates occurring in some jurisdictions due to generation construction cost overruns; they demanded access to the cheaper power being sold in other areas or produced by efficient and quickly built plants burning now-cheap natural gas. Rates in neighboring states sometimes differed by a factor of two. In some areas, the difference between the low marginal costs of producing power by gas and the high embedded cost rates of vertically integrated supply created a strong incentive for consumers to access cheaper power.

The 1990s witnessed the debunking of a long-standing myth that the electricity industry is a natural monopoly, with competition being introduced in many jurisdictions in two forms. One was PURPA-style competition among independent generators to sell to utilities at a regulated avoided cost-based price. To the extent that these avoided costs accurately reflected the cost of utility-based supply, this type of competition reduced the cost of supply. However, regulators often overestimated avoided costs, which sometimes resulted in overpayment and expansion of uneconomical supply, which, in turn, inflated electric rates. This increased pressure for a second type of competition: access by consumers to cheaper generation, facilitated by marketers.

At first, this access was through wheeling arrangements. But to spread widely the benefits of access, many states turned to a restructuring framework in which distribution, transmission, and generation were unbundled. Generation assets were then to be spun off to independent supply companies that would then compete to sell power to consumers, using the grid as a common carrier. This second type of competition also did not necessarily decrease the overall societal costs of power. It has been argued that it was primarily an effort by some customers to shift the fixed costs of stranded assets (such as uneconomical PURPA contracts or utility generation whose construction costs were underestimated) either to other customers or investors (Borenstein and Bushnell, 2015). Today, similar efforts by consumers to shift costs to other parties by avoiding average cost–based rates are continuing to drive

https://www.nxtbook.com/nxtbooks/pes/powerenergy_010219/index.php#/p/Cover4 2/9

> Today, similar efforts by consumers to shift costs to other parties by avoiding average cost–based rates are continuing to drive radical changes in the industry.

radical changes in the industry—but in a very different way, as we explain later.

Genesis of the Second Wave

In the 1990s, the enthusiasm of many states for restructuring was fed by overestimated savings from competition; optimistic projections of the competitiveness of new, cleaner technology; and the prospect that the some of the benefits of competition could be tapped for popular programs subsidizing energy efficiency and renewable technology. The belief that restructuring was a cornucopia to benefit everyone was epitomized by the unanimous vote of the California Assembly in favor of Assembly Bill 1890 in 1996, the Electricity Restructuring Act. Other states quickly followed suit. Within the next few years, nearly half of the states had passed laws or instituted regulations to restructure the power industry. Many of those states intended to allow retail access for all customers, optimistically anticipating wide participation and dramatic cost savings. These state actions were followed in many cases, most notably in California, by widespread voluntary or forced divestment of generation assets by the former monopoly utilities.

Unfortunately, if such a dramatic change to an industry is favored by everyone, at least some people must be overly optimistic about the benefits they will receive. Any such change, even if it improves efficiency and lowers prices, will inevitably produce some losers as well as winners, which, in fact, became apparent later. Furthermore, as many industry writers warned at the time, the technical intricacies of power systems meant that creating a market for power would not be nearly as easy as for much simpler systems. The inspiration provided by the successful creation or restructuring of markets for simple commodities, such as the 1990 Clean Air Act's continent-scale pollution-trading scheme, was, in part, misleading.

At the same time that states were acting, the Federal Energy Regulatory Commission (FERC) also moved the restructuring agenda forward. In April 1996, it adopted Order 888, which was intended to "remove impediments to competition in the wholesale bulk power marketplace and to bring more efficient, lower cost power to the Nation's electricity consumers." All public utilities that owned assets transmitting electricity in interstate commerce had to allow market parties to use their facilities on a nondiscriminatory basis, based on published transmission tariffs. Making the high-voltage grid available to all comers would, it was hoped, facilitate the states' retail markets and give access to the independent generators that the states facilitated through their unbundling and divestment actions. In December 1999, FERC issued Order 2000, which explicitly encourages the establishment of regional transmission organizations (RTOs) to facilitate markets and further the goals of Order 888. Some of those RTOs have since become full-fledged independent system operators (ISOs) that operate day-ahead and real-time electricity spot markets, although many parts of the country honor Order 888 simply by complying with rules requiring nondiscriminatory access to their transmission facilities.

The First Wave Disappoints: Genesis of the Second Wave

The California Crisis

The 2000–2001 California crisis, however, stopped state restructuring efforts in their tracks. The causes of the crisis have been chronicled and debated extensively in this magazine. We believe that the problems arose, first, from an exercise of market power right out of an Economics 101 class—the withdrawal of supply to push up prices. This exercise was facilitated by supply problems, such as drought, a natural gas compressor explosion, and kelp sucked into a nuclear station's cooling system and was also enabled by state regulators who actively discouraged forward contracting.

Second, the market design also disregarded important physical characteristics of electricity. These naïve designs included the definition of uniform pricing zones in the day-ahead market that ignored local transmission congestion. As a result, operators had to scramble to resolve congestion in real time, endangering reliability and exacerbating local market power.

Another issue was the creation of isolated day-ahead and real-time markets for energy and operating reserves that overlooked their strong interconnections. Enron and other market parties resorted to disguising actions or outright lying to arbitrage between those markets. Such arbitrage is, in fact, an essential function in all other markets and arguably improved rather than harmed power market functioning at the time.

There was also the failure to promote forward contracting, which enabled any exercise of market power to immediately benefit the whole portfolio of a company's generators. The consequences of poor market designs and market power were unambiguous: rolling blackouts, a threefold increase in wholesale power costs, and loss of public confidence in the ability of restructuring to deliver lower costs. Since the crisis, no additional states have joined the original 23 to allow for retail choice in their electricity markets, and, of those 23, eight have suspended such choice. Of the states that originally encouraged retail markets, only Texas has a vibrant and active retail markets for all consumers: industrial, commercial, and residential.

Despite the crisis, FERC gamely pressed on. After an unsuccessful attempt to develop a standard market design to impose on all restructured wholesale markets, a voluntary wholesale market "platform" was promoted, with elements that FERC believed would prevent California-style problems from occurring in the future. These included location-specific (nodal) prices that reflected consideration of the network's actual constraints, before-the-market identification and mitigation of bids that appear to attempt to exercise market power, forward financial contracts and resource-adequacy markets, financial transmission rights (FTRs), cooptimized spot energy and operating reserves markets, financial arbitrage between day-ahead and real-time markets, and a hybrid market that operates like a so-called power pool company (POOLCO) but also facilitates bilateral transactions.

A POOLCO in a particular region is defined as a single market operator that would run an auction to buy power from generators and then resell it to consumers or load-serving entities. POOLCOs can also enable bilateral deals to be made between generators and consumers or load-serving entities by moving the power from the former to the latter and charging for the short-term use of the network based on the differences in the energy spot prices it calculates at the locations of sellers and buyers. This allows trading to take place, while accounting for all of the physical constraints in the network that must be satisfied.

Lessons Learned from the First Wave

To summarize, the first wave of restructuring was driven by a global move toward privatization of infrastructure industries, vertical and horizontal unbundling, and implementation of competitive wholesale markets as means for achieving economic efficiency in investment, production, and consumption. We now describe several of the lessons learned from the first wave of restructuring and the California crisis, lessons that are, more or less, embodied in FERC's platform and in the market redesigns actively promoted in PJM Interconnection, California, and the other U.S. ISOs.

One lesson concerned how to unbundle and then distribute the resulting economic benefits and costs. Not everyone automatically becomes better off. The initial main challenges were the divestiture of assets and recovery of the stranded costs resulting from the regulatory compact. Some thought that electricity under deregulation would be much cheaper, which would, therefore, significantly depreciate the value of the divested assets. In California, the Competitive Transition Charge mechanism was designed to make the shareholders of the investor-owned utilities whole. But the arrangement was asymmetric and did not protect customers against windfall profits for the asset buyers due to unexpected events and the exercise of market power. No one thought of protecting consumers' stranded benefits (earned through depreciation payments embedded in the regulated rates) through vesting call options (imposing bid caps) on divested plants. Such vesting contracts were imposed, e.g., in the earlier U.K. divestiture. The managers of future restructuring, which may eventually come to the northwestern and southeastern parts of the United States that have resisted it, should consider smoothing the transition through such contracts.

Another lesson was the need for consumers to forward-contract for supply to prevent the exercise of market power. The California regulators discouraged long-term contracting between generators and load-serving entities, believing that clearing all electricity in the day-ahead spot market would eliminate the need for regulatory intervention in price setting. After the crisis, California regulators reversed course and instead mandated month- and year-ahead forward contracting and showings of adequate physical resources. FERC, in turn, has encouraged, but not mandated, so-called resource-adequacy mechanisms in part to give more assurance that enough supply and demand resources would be available to prevent shortages. After the crisis, these mechanisms helped assure nervous regulators that another California blowup would not happen and provided a source for missing money for resource investors who would not earn adequate returns from price-capped energy markets alone.

The adopted approaches, however, differ in different areas. PJM and ISO New England have capacity markets in which the ISOs procure forward capacity (subject to specific must-offer obligations) through annual auctions that allow participation by existing and planned capacity as well as demand-side resources willing to curtail load. Texas, however, has opted for an energy-only market, with spot prices allowed to reach US$9,000/MWh. It has also complemented the raising of price caps with an administrative operating reserve demand curve that imposes a price adder on energy to reflect reserve depletion, so that energy prices better reflect scarcity rents in times of shortage.

A third lesson was learned when both state regulators and FERC underestimated the market power and other abuses that could take advantage of inconsistencies between the commercial model (e.g., uniform price zones and separate energy and ancillary service markets) and the technical reality. FERC authorization of market-based wholesale rates was based on standard Department of Justice criteria—e.g., the Herfindahl–Hirschman concentration index—that proved to be inadequate for the electric power industry. The commercial models initially implemented in California and Texas emphasized trading simplicity, as is still the case in Europe. The California crisis and widespread gaming (particularly the so-called "dec game") in PJM, California, and Texas demonstrated the need for consistency between the commercial and technical models, which eventually led to further reforms of these markets, such as California's market restructuring and technical upgrade and the Texas nodal market. (In the dec game, a generator in a location with excess power would schedule a power sale into a zonal day-ahead market at a regional high price that disregards congestion and then buy the power back in real time at a lower price that reflects local congestion.)

In particular, nodal markets, whose locational marginal prices would be determined through uniform price-clearing auctions, were instituted in both the day-ahead and real-time markets. Initially, real-time markets were hourly, but to address the need for greater flexibility in response to renewable output variations, FERC Order 764 reduced the settlement intervals to 15-min markets, with reoptimization of dispatch every 5 min. Contrary to fears that having separate prices at each bus would increase the ability to manipulate prices, the markets diminished opportunities to exercise market power by making constraints transparent and managing congestion day ahead. However, these measures are not enough to eliminate market power in load pockets with limited supply and transmission bottlenecks.

From the California crisis, we learned that there was a need for active market monitoring together with automated detection and mitigation of inflated supply bids that attempt to exercise market power in local markets. Trading energy at nodal prices has provided accurate price signals reflecting congestion that facilitate efficiency in production, consumption, and investment. However, it quickly became apparent that the variability and uncertainty of such locational prices exposes market participants to high risks. Thus, another lesson was that there was a need for financial hedges for these risks. Bill Hogan of Harvard complemented Schweppe's vision of nodal energy markets with a proposal that ISOs create hedging and property rights instruments in the form of FTRs. These are financial instruments that entitle their holders to the net gain resulting from nodal price difference between two specific nodes. They allow energy market participants and transmission owners to hedge the locational price risk associated with energy trading and provide means for monetizing some of the property rights resulting from the ownership of transmission assets that restrict power flows. Markets for FTRs, with periodic auctions, are managed by all ISOs as a complement to the model energy markets (Hogan, 1992).

Everyone also learned from the California debacle that markets should not be arbitrarily isolated and that arbitrage is important for improving efficiency, providing information, and making it harder to exercise market power. One response is to design markets so that energy and ancillary service interdependencies are recognized; this is now done by cooptimizing those markets, committing resources to provide whichever commodity is most valuable to the system and paying in a way that also makes that schedule the most profitable course for individual suppliers. Another response is the two-settlement energy market design, with virtual bidding, which is now standard in U.S. organized markets. Thus, most energy trading is contracted in forward bilateral and over-the-counter markets, complemented by long-run financial contracts. The ISOs settle day-ahead and spot energy transactions through a two-settlement approach where most energy is traded hourly for the next 24 h with financially binding quantities and prices determined by a security-constrained, bid-based economic dispatch, implemented through unit commitment optimization employing mixed-integer programming.

Deviations from the day-ahead hourly commitments are reoptimized every 5 min based on rebidding every 15 min. Virtual bidding allows arbitrage by both energy traders and speculators and results in a closer convergence of day-ahead and real-time prices. It also eliminates the incentive to disguise arbitrage through implicit virtual bidding, in which physical energy offers or bids to buy are deferred until real time to manipulate day-ahead prices. Both the supply and demand sides engaged in implicit virtual bidding during the California crisis. Figure 1 illustrates a typical day-ahead and real-time dispatch versus the hourly day-ahead load forecast and actual load.

The Second Wave: Focus on Economic Efficiency and Incremental Improvements

What was learned in the decade after the California crisis is that market design is not a destination but a journey. As supply and demand technologies improve, computational capabilities are enhanced, and more is learned about the functioning and needs of markets, further changes are proposed and implemented. In recent years, these changes have been primarily incremental in nature (which we call the second wave), but as we will discuss in the next section, more dramatic reforms may lie ahead in response to the smart grid, renewable energy, and distributed resources revolutions.

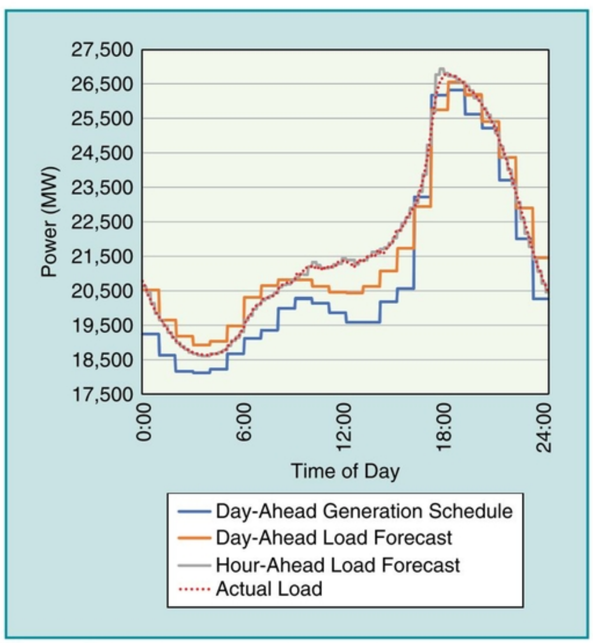

figure 1. A typical two-settlement dispatch, from 8 January 2018. (Data from California ISO.)

One fundamental challenge is the divergence between the economist's ideal of a convex, smooth supply cost function and the reality of lumpy commitments, minimum output constraints, and nonconvex heat rate curves. The U.S. market designs allow generation resources to provide economic offers in the day-ahead and real-time markets in the form of various combinations of an energy supply function, start-up cost, and no-load cost. Such offers are cleared using unit commitment optimization mixed-integer programming software. Alternatively, generation resources can submit physical bilateral transactions that are scheduled outside the optimization as price-taking transactions.

One of the consequences of the nonconvex cost structure of generation is that generation units may not cover their lumpy costs through energy payments they receive over a 24-h period. But because of another widely adopted market feature, generators who participate in the market by submitting economic offers receive make-whole payments, covered by an uplift known as *bid cost recovery* (*BCR*) payments that ensure no losses over each 24-h period. The BCR payments provide an incentive for generators to submit economic offers rather than self-schedule.

The rules for bidding and paying for lumpy costs have remained an important item on the market reform agenda, with ongoing dissatisfaction over the basic BCR model. To minimize uplifts and recover as much of the cost as possible through energy charges, an approach known variously as *extended locational marginal pricing* and *convex hull pricing* has been partially implemented at the Mid-Continent ISO and is being considered by other ISOs.

Another area of incremental improvement has been the geographic expansion of markets to take advantage of complementary fuel mixes in neighboring regions and short-term diversity in load and, increasingly, renewable output. PJM's expansion westward to Chicago immediately resulted in increased west-to-east power flows and in fuel cost savings of over US$100 million per year. The integration of the day-ahead markets in Europe has achieved similar results.

More recently, the desire for more efficiency in real-time balancing in western North America has led to the rapid expansion of the California ISO's energy imbalance market (EIM) involving eight U.S. states and even the Canadian province of British Columbia. A large motivator is the negative real-time prices that often occur midday in the California market, which means that neighboring states would be paid to take California's surplus. The EIM eliminates the hurdle rate of transmission access and other charges that hinder trade between neighboring balancing areas. Unfortunately, governance issues have prevented the California ISO from expanding this model to the day-ahead market. Sensing an opportunity, the Southwest Power Pool/Mountain West partnership is competing to enlist balancing areas that have not yet joined the EIM. It also seeks to expand day-ahead markets, perhaps including closer coordination of the eastern and western interconnections through expanded dc links.

Another trend made possible by computational software and hardware improvements is the explicit consideration of postcontingency corrective actions in market software. Previously, such risks were handled in the markets by preventive prepositioning of supply, through zonal operating reserve requirements, *N*-1 constraints on dispatch, or minimum online constraints. The hope was that this excess supply would be sufficient to allow the system to securely manage unexpected equipment outages or fluctuations in net loads. For some time, academics have pointed out that if postcontingency redispatch could be modeled and optimized, then the cost of prepositioning the generation could be reduced. This is because the software would automatically fine-tune the location and quantity of reserves that are acquired.

The California ISO is implementing corrective action optimization for certain contingencies by explicitly representing how dispatches and commitments could be adjusted optimally after an event. Similar approaches could be used to optimize reserves to cover, e.g., ramp events, avoiding the need for highly conservative and blunt zonal or system requirements for flexible ramping products. Ultimately, this could lead to full-blown stochastic optimization for unit commitment, in which a large number of possible contingency, load, and supply scenarios could be considered when optimizing which units to put online and which fast-start units to hold in reserve.

The incremental improvements associated with the second wave have resulted in or will soon yield significant operating cost savings. Not all parties, however, are happy with this trend toward making market software more complex. Some complain that price drivers are not transparent and perhaps have contributed to decreases in trading activity. An evolution toward stochastic market clearing could accelerate that trend. Some observers call for a reversal of this trend, e.g., toward simplifying supply offers by eliminating lumpy cost and minimum output bidding. But this discussion has received far less attention than other looming issues that the industry in general and market operations in particular are facing.

The Third Wave: Making Way for Demand-Side, Renewable, and Distributed Resources

The third wave of market and industry reform, which is largely still a work in progress, is associated with the rapid penetration of renewables and distributed energy resources (DERs) into the electricity generation mix and the deployment of smart grid technologies, such as smart metering, phasor measurement units, and storage (Sioshansi, 2016). Also important are changes in consumption patterns due to demand response, the electrification of the transportation sector, and increased customer participation in electricity production (so-called prosumage). As we discussed earlier, the first two waves were largely motivated by the move toward vertical and horizontal unbundling and privatization of the electricity infrastructure and by increased efficiency of investment and operation through competition. In contrast, the third wave is largely driven by an environmental agenda for the decarbonization of electricity generation to reduce global warming, a social movement toward more exercise of customer choice and democratization of the energy supply, and rapid technological innovation in supply, storage, metering, and control in the energy area.

Political forces advocating for these trends have championed policies that subsidize technological innovation on ideological grounds rather than economic efficiency. This tilting of the playing field is counter to the fundamental principles of market design. The result is perverse incentives and rent-seeking behavior on the supply and demand sides alike and proposals for increasingly complicated market mechanisms that attempt to mitigate the technical challenges and economic distortions resulting from those incentives.

In the rest of this article, we consider some of the implications for the third wave of market design of the accelerating penetration of renewable and distributed sources. For example, policies such as net metering have become pervasive in the United States and abroad. These policies compensate rooftop photovoltaic (PV) solar production at the retail rate, even when household production of solar energy exceeds consumption. To illustrate, in California, net metering payments under Pacific Gas and Electric's retail tariff E6 can reach US$0.45/kWh at peak hours, while the wholesale energy cost may be as low as US$0.04/KWh. Such rates fail to recognize that the high volumetric retail rates were designed to distribute the infrastructure cost among retail customers in an equitable manner so that consumers that use more energy pay a markup reflecting a higher share of the infrastructure cost. Under the net metering paradigm, however, such high-end consumers can offset their payments by installing solar panels and shifting their share of the infrastructure cost to low-end consumers. The distortion created by net metering has resulted in rapid expansion of rooftop PV systems in states such as California, Nevada, and Hawaii, even though their costs are roughly twice that of grid-scale solar.

This expansion, in turn, has led to a legislative rebellion that discontinued the original net metering policy in Nevada and Hawaii and produced a close three to two vote in the California Public Utility Commission that revised net metering practices and rates. Hawaii has opted for a major reduction in payments for the excess production of rooftop PV systems above customers' self-use and an increase in fixed connection charges for new rooftop solar installations. Nevada went even further along that path by retroactively changing the terms originally offered to customers with rooftop solar installations. These moves were fought in the political arena by the rooftop solar power industry.

While net metering may distort economic incentives in favor of rooftop solar installations, it should be noted that centralized PV solar power is becoming increasingly competitive, and the U.S. Department of Energy has recently announced that centralized PV solar installations have reached long-term costs of US$0.06/kWh. Consistent with that, in Israel, the Public Utility Authority has recently signed a 20-year supply contract for PV solar power at US$0.057/kWh, while in Chile's Atacama Desert, contracts for PV solar at US$0.033/kWh were recently signed. Such supply for renewable energy would easily compete in current electricity markets without the need for perverse incentives inducing inefficient rent chasing, especially if a carbon tax or a cap-and-trade policy is implemented.

Wind power is another major renewable energy source that could be competitive without targeted subsidies if appropriate carbon pricing were introduced. Unfortunately, in many systems, wind power has been implicitly subsidized through feed-in tariffs and must-take policies imposed on system operators. Such policies are often supported by the false premises that renewable energy is free, is valuable for its own sake, and should not be spilled except for emergencies or overgeneration situations. These suppositions do not account for many of the complexities associated with providing reliable power considered in security-constrained, economical dispatch. It can be easily demonstrated that the strategic curtailment of the renewables supply (which is technically feasible) enables the more efficient and even less polluting dispatch of conventional slow-ramping resources. Hence, in some systems, such as the California ISO, negative bids (down to –US$150/MWh), and thus prices, have been incorporated into the market design to negate some of the perverse production subsidies. Further incentives for the curtailment or withholding of renewable energy can be used to correct the distortions resulting from suboptimal deployment of conventional standby resources.

In Europe, and particularly in Germany, the rapid penetration of wind generation and solar power through perverse economic incentives has resulted in the virtual collapse of the wholesale energy market. As shown in Figure 2, such a collapse in wholesale energy prices resulted in a missing money phenomenon that has motivated the capacity remuneration of conventional power plants. Ironically, the abundance of renewable power has also resulted in a sharp drop in carbon prices and led to the increased use of lignite, which negates much of the environmental benefit provided by renewables.

From a market design perspective, the rapid penetration of renewables and DERs has created several major challenges.

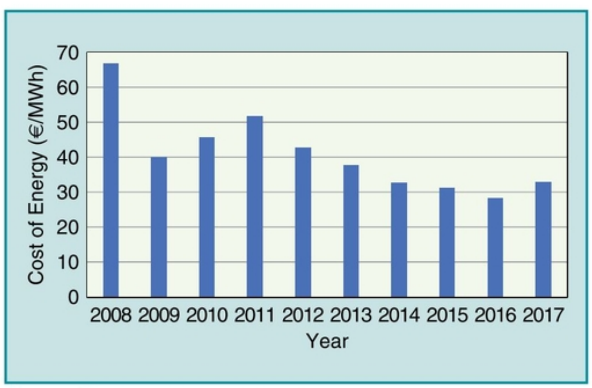

figure 2. The collapse of the German wholesale market. (Data from energy-charts.de.)

The inherent systematic variability and uncertainty of renewable resources results in extreme ramping and a need for flexibility. The famous California duck curve, shown in Figure 3, illustrates the impact of high solar penetration on the need for fast-ramping resources in the early morning and late afternoon and the overgeneration potential that would require curtailing renewable resources. To address these problems, the California Public Utility Commission has modified its resource adequacy (RA) mechanism to require about a third of the capacity that load-serving entities must show for achieving their obligation needs to meet flexibility criteria. Flexible resources must be economically dispatchable and able to be ramped continuously over a 3-h period. Unlike regular RA resources, which must offer obligations by being dispatched or available for dispatch, flexible RA must be offered at a price that enables the ISO to dispatch them economically.

For several reasons, however, the RA mechanism is a blunt tool to encourage flexibility. First, is it 3-h, 1-h, or 5-min ramps that must be managed? Different resources are best for each. It also matters whether the ramps are predictable or uncertain. Second, how does one compare, for instance, the flexibility contributed by combustion turbines that are limited to one start a day, storage with 15 min of energy, demand response that can be called upon only a handful of times per month, and plants with different ramp rates? This suggests that, rather than develop a generic definition of flexible capacity, it would be better to reform energy markets to reward capacity that is available precisely when needed and can ramp up quickly when prices spike and then disappear when they collapse.

But reforming energy and ancillary services markets to provide more appropriate rewards for flexibility is not necessarily simple either. Flexibility of energy resources is essential for the integration of renewables and DERs at various time scales, and hence such flexibility must be priced in the various market horizons so that it provides the appropriate price signals for investment and operation of resources. This is accomplished by creating revenue streams and market products that enable the remuneration of resources that provide flexibility. Such flexibility can be mobilized through shorter settlement intervals, demand-side management, the flexible use of transmission assets (e.g., transmission switching and dynamic ratings of thermal limits), and flexible ramping products. The value/cost of flexibility must be reflected in the payments to renewable and distributed resources that impose such costs on the system, and resources that help resolve issues associated with flexibility should be remunerated.

For example, California ISO has introduced a new flexible ramping product (sometimes called *flexiramp*) that enables remuneration of resources that are being held in reserve out of merit order instead of being dispatched to produce energy which they offer below the market clearing price. Such reserves differ from regulation-type reserves, which

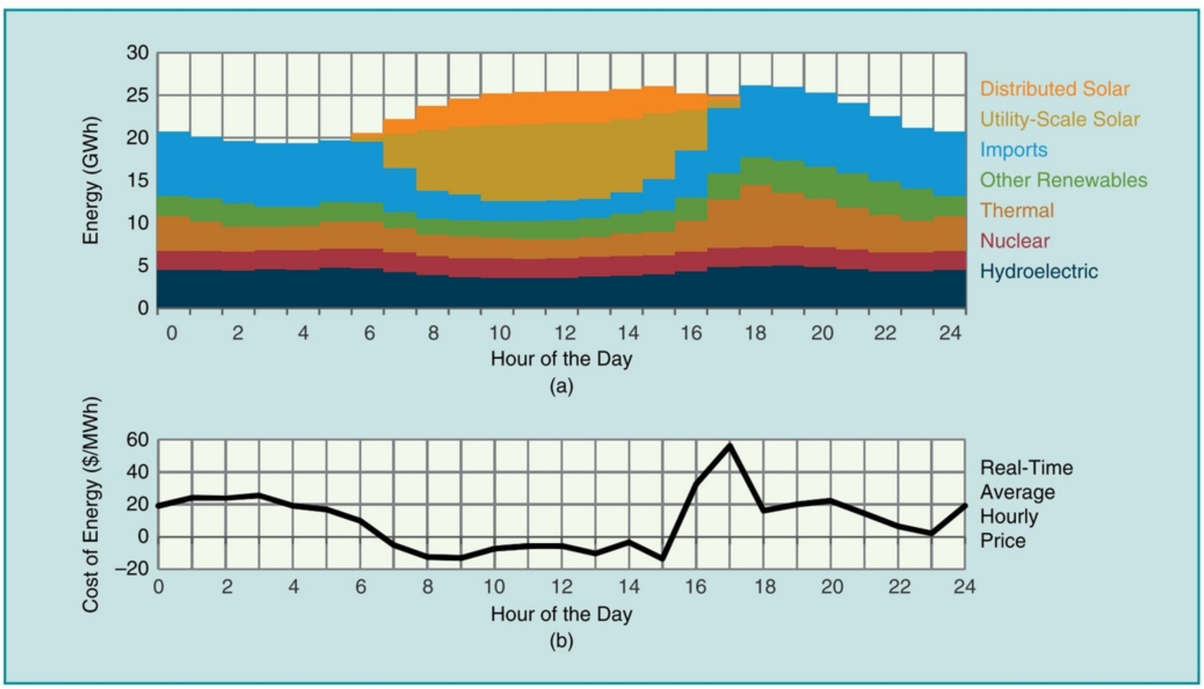

figure 3. An illustration of (a) the California ISO net generation duck curve and (b) negative wholesale prices, both for 11 March 2017. (Note: Distributed solar generation is estimated based on December 2016 installed net-metered capacity as reported in form EIA-826, Monthly Electric Utility Sales and Revenue Report with State Distributions, U.S. Energy Information Administration.)

are a capacity-based ancillary service product used for load following within each 5-min interval. In contrast, flexiramp serves to assure the availability of energy and regulation in future intervals. Hence, these flexiramp reserves are remunerated for lost opportunity costs, which is the profit that is forgone when capacity is held back out of merit rather than dispatched (i.e., opportunity cost = clearing price – bid).

Besides the growth of renewables, the other dramatic change in the industry is the proliferation of many small DERs on both the demand and supply sides. This reflects a general trend toward diverse participation in the production of energy and the above-described need to address the many new challenges posed by the introduction of renewables. Such proliferation is largely enabled by innovation in the smart grid technologies area. There are many ways to integrate such resources and organize market platforms that will facilitate the procurement of DERs, enable peer-to-peer transactions at the distribution level, and create new products. These changes will redefine the role of future utilities to support their ability to assure retail supply reliability in a world where their volumetric revenue base is shrinking because of increased self-supply of energy behind the meter.

One market design choice is whether to make the market ISO-centric, where DERs are integrated into the system through distribution system operators and third-party aggregators who participate in the ISO wholesale market. Alternatively, the market organization could be distribution-level–centric, with the ISO performing limited coordination functions among distribution system operators that accomplish most of the DER integration through local distribution system market platforms. The California ISO and PJM have opted for the first through-market initiatives that carve a role for DER aggregators (in California) and curtailment service providers (PJM) that behave as virtual power plants that can participate as regular plants in the ISO wholesale market. Such participation broadens the scope of the conventional ISOs and adapts them to the brave new world of renewables and distributed resources. New York State, on the other hand, has been promoting its Reform the Energy Vision initiative, which emphasizes customer choice and active participation by DERs.

Conclusions

From Thomas Edison's Pearl Street Station (and his earlier Holborn Viaduct plant in London) to Samuel Insull's regulated utility and the vertically integrated monopolies of the late 1900s, the power industry saw tremendous innovation in technology and institutions during its first century. The last three decades have seen further tremendous changes, with environmental concerns coming to the fore and renewable generation now dominating new capacity additions. New computation and communications technologies have facilitated a revolution in markets, allowing much bigger geographic coverage and efficiently coordinating an array of energy, ancillary service, and transmission products. This growth in the scope of markets lowers costs by increasing competition and taking advantage of diversity of load and generation patterns across large regions. Ironically, while the scale of markets has greatly increased, at the same time the scale of who can compete has become much smaller, so that household-scale technologies are now revolutionizing the market.

Market designs have not kept up with these changes. This means that we have not realized the full economic and environmental benefits of new technologies. Present challenges for market designers include further expansion of the geographic scope of markets; reconciling the wildly inconsistent price signals in retail and wholesale power prices; incentivizing an optimal mix, location, and scale of flexible generation, storage, and demand-side resources to firm up renewable generation; and avoiding confusing means for ends by keeping a firm eye on the goals we want to accomplish with renewable and other policies. It is lower system costs, greater reliability, and lower pollution we want, not particular technologies; the best means to achieve those goals are likely to surprise us. Therefore, markets must be designed to allow the best solutions to emerge and succeed without policy makers imposing their prejudices about what they should look like.

For Further Reading

S. Borenstein and J. Bushnell, "The US electricity industry after 20 years of restructuring," *Annu. Rev. Economics*, vol. 7, no. 1, pp. 437–463, 2015.

W. W. Hogan, "Contract markets for electric power transmission," *J. Regulatory Econ.*, vol. 4, no. 3, pp. 211–242, 1992.

P. L. Joskow and R. Schmalensee, *Markets for Power: An Analysis of Electric Utility Deregulation*. Cambridge, MA: MIT Press, 1983.

E. T. Mansur and M. W. White. (2012, Jan. 13). Market organization and efficiency in electricity markets. Dartmouth College, Hanover, NH. [Online]. Available: www.dartmouth.edu/~mansur/papers/mansur_white_pjmaep.pdf

F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn, *Spot Pricing of Electricity*. Norwell, MA: Kluwer, 1988.

F. P. Sioshansi, Ed., *Future of Utilities, Utilities of the Future: How Technological Innovations in Distributed Energy Resources Will Reshape the Electric Power Sector*. New York: Academic, 2016.

S. Stoft, "Economic feedback: Working with the insatiable profit motive," *IEEE Power Energy Mag.*, vol. 99, no. 3, pp. 14–16, 2003.

Biographies

Benjamin F. Hobbs is with Johns Hopkins University, Baltimore, Maryland.

Shmuel S. Oren is with the University of California, Berkeley.