ENSURING ADEQUATE GENERATION SUPPLY IN COMPETITIVE ELECTRICITY MARKETS

Dr. Alex Papalexopoulos
Fellow, IEEE

ECCO International, Inc.
San Francisco, CA
alexp@eccointl.com

1. INTRODUCTION

A key problem policy makers and market designers are concerned with in a competitive energy marketplace is how to ensure sufficient generation supply to meet demand and ensure reliability in the long term. Historically, vertically-integrated utilities maintained a 15 to 20 percent reserve margin as part of their obligation to serve, often required by their local state commissions. Also in the past utilities undertook integrated resource planning that coordinated investments in generation and transmission to meet predicted load growth and ensure reliability. This solution, however, is not applicable in the new environment for the following reasons. Independent power producers now undertake most investments in generation. Few Load Serving Entities (LSEs) are self-sufficient in generation, relying on significant purchases from other wholesale producers. Furthermore, electricity markets are regional and extend beyond state boundaries. Price-responsive demand could potentially be an integral part of the solution to this problem. However, due to technological, investment and regulatory obstacles, such solutions cannot be put in place quickly and other mechanisms are needed to ensure adequate generation supply.

Before restructuring, a vertically integrated utility typically owned and operated sufficient generation capacity and could each day commit and schedule sufficient units to provide reserves for the next day. This practice is performed now by the ISO using a day-ahead co-optimization of units' schedules for energy generation and reserve capacity, or a separate day-ahead market for reserves, for units not obligated by bilateral contracts. The commercial implication is that the RTO pays LMP prices that are volatile.

A common view is that LMP is the appropriate market mechanism to secure sufficient generation capacity and ensure reliability. In fact, however, these volatile short-term price signals are imperfect indicators for long-term investments. One adverse factor is that an investment on a large scale eliminates the profit opportunities by depressing the LMP price or eliminating the LMP differential between various locations. Further, transmission investments typically have external effects throughout the grid, so an investor who expands the capacity of one line may not capture all the resulting benefits.

The concept of electric system reliability encompasses two distinct aspects, "security" and "resource adequacy." Security addresses the systems capability of sustaining short-term disturbances. Resource adequacy defines the capability of the electricity system to meet predicted demands. In economic terms, system security is a "public good" reflecting a global characteristic of the power system. Like other public goods such as national defense or clean air, it is not possible to exclude those that refuse to share the cost from enjoying the benefits, i.e., free ridership is possible. Hence system security requires central oversight and mandatory participation.

In the case of resource adequacy the dividing lines are not obvious. The concept of "obligation to serve" any load with some predetermined probability is incompatible with the notion of a competitive market in which available supply depends on the price customers are willing to pay for it. In such a setting it is meaningless to talk about the probability of not being able to meet load without specifying the price customers or LSEs are willing to pay for the provided energy. Hence the "obligation to serve" concept must give way to an "obligation to serve at a price."

In a perfect energy market, where prices are allowed to fluctuate, consumers would be able to insure themselves against price spikes through fixed-price contracts or hedging instruments in the same way that they protect themselves against interest rate fluctuations by assuming fixed-rate loans. The market prices for such hedges or the premium for fixed-price contracts provide market signals for generators who may consider investing in new generation capacity.

Unfortunately, the ideal economic scenario outlined above is predicated on a host of assumptions about regulatory, political and technological considerations that would allow customers to respond to prices in real time and would allow the distribution company to selectively curtail power to consumers that choose to be curtailed when the price exceeds certain levels. Nevertheless it is still useful to think of generation adequacy in terms of price insurance that is aimed at protecting the public against extreme prices due to shortages, rather than as a reliability issue.

The alternative view, which is more in line with traditional engineering concepts, is to think of generation adequacy in terms of service reliability that is treated as a separate product from energy and is provided through generation capacity, i.e., “steel in the ground”. Under this approach LSEs have capacity obligations that they can meet through capacity markets. The procured capacity typically entails an obligation to be available to produce energy but no price limit at which the energy must be offered other than a global bid cap.

Criteria such as 15% reserve margins, or one day of outage in ten years, are based on loss-of-load probability calculations and an implicit assumption about the value of lost load [VOLL]. In a market-based setting the market determines the VOLL through market mechanisms that bring together supply and demand. Reserve margins alone, however, do not necessarily translate into an ability to meet demand. Persistent shortages of natural gas, gas transportation capacity and emission allowances could render “steel in the ground” all but useless. In the Eastern Interconnection, transmission failures and frozen coal piles are far more likely to result in widespread electric service interruptions than would be evident from the lack of generating capacity. Furthermore, reserve criteria may be sufficient in the capacity-limited East, where hydro accounts for 5% of the energy generated; they may woefully inadequate, however, in the energy-limited West, where hydro has contributed as much as 40% of energy generated (in 1997) and as little as 22 % (in 2001).

This paper summarizes the available options for solving the problem of ensuring adequate generation supply and proposes a methodology that offers some advantages over other approaches currently in operations. Section II presents the main options currently available for ensuring adequate generation supply and reliability. Section III presents an overview of the approaches currently implemented at the Eastern ISOs in the US. Section IV proposes a methodology that offers some advantages over other approaches currently in operations. The main conclusions are summarized in section V.

2. MAIN OPTIONS FOR ENSURING GENERATION SUPPLY

There are two extreme options for ensuring adequate generation supply and maintain reliability. The energy-only market option and the regulatory authority and/or the ISO based option in which a unilateral action to build generation is taken by the state and/or the ISO. Between these two extreme options several mechanisms exist to encourage new capacity construction.

2.1. Energy-Only Market Option:

Under this option no formal capacity assurance mechanism is put in place to ensure sufficient generation supply. Energy and ancillary services prices in the spot and forward markets fluctuate and when they are high enough they justify new investments. There are several energy only electricity markets around the world, including the original California market, the Australian Victoria pool and the Nordpool. Theoretically, when demand exceeds supply scarcity rents, i.e., the difference between the price of the most expensive unit online and the demand curve will cover the capacity cost of these generators. Furthermore, they will induce sufficient demand response so that available supply can meet the remaining demand. A shortage of capacity will have the effect of increased scarcity rents, and increased prices, thus increasing the probability for investment. Excess capacity, on the other hand, will drive prices to marginal cost.

The high price volatility in an energy-only market has major political implications but it is not the only problem regulators and politicians are concerned with. In this type of market no resource planning is in place. No single entity has the clear authority to project resource shortages and make the necessary arrangement for adequate reserve to be in place. Private investors will only respond to short-term or spot market energy prices. Little investment will take place in low price years, causing shortages to develop. That is what happened in California in the late 1990s. Although theoretically, extremely high price spikes can be sustained, new capacity cannot come overnight, thus creating an unacceptable transfer of wealth from consumers to generators.

Due to the lack of planning and coordination, when capacity does come on-line, there is generally overbuilding, which can lead to depressed market prices and deter new investments. The resulting cycle of shortages and high prices will repeat again.and this boom/bust circle can continue without end. Some argue this cycle will induce more investment in price responsive demand that will help moderate these cycles. This may be true, but during the bust period (shortages), system reliability will suffer. Thus, the energy-only market model will ultimately fail to ensure system reliability and will cause exposure to huge price volatility and market power.

Furthermore, in the absence of a reliability, or a capacity price mechanism the consumption (or generation) of energy creates external reliability costs (or benefits) that are not explicitly priced. Specifically, market participants impose a reliability cost without having to pay for it, and generators confer a reliability benefit without being compensated for it.

2.2. Unilateral Action by the State/ISO:

The other extreme to the energy-only market paradigm is based on the framework where the Government/State is financing capacity investments. In many cases, this option is being contemplated as a replacement for the energy-only market option. While it may be a temporary supplement of private investment, it may be inefficient as the main mechanism to ensure sufficient reserve margin. It returns to central planning, and could adversely affect innovation and efficiency gains from involvement in the private sector. With proper planning, system reliability should be protected, but with increased cost to the state and ratepayers due to a decrease in private investment and potential loss of efficiencies.

Between these two extreme alternatives there are several other options to encourage new capacity construction. There are briefly presented next.

2.3. Administrative Payments for Capacity:

With this option a regulatory mechanism of payments is established to encourage capacity investments. Argentina, Colombia and Spain have implemented such mechanism to ensure sufficient generation supply. Generators are paid based on availability, technology, LOLP, VOLL to incent investment and availability. In theory this mechanism can work, but in practice it creates several distortions. These distortions are related to the level of payments to generators and to the perception the customers have that receive no value for their money.

England & Wales had originally implemented a variant of this approach by adding a reliability adder on the pool prices. The reliability adder was based on the product LOLP x (VOLL-SMP). One way to improve this approach is to adjust the capacity adder according to the energy price bid by the generator. Thus, dispatched generators would receive an option premium based on the hourly SMP serving as strike price while generators whose bids exceed the SMP should be paid a call option premium according to their energy bid serving as strike price.

Another key problem of the administrative capacity payment is that it is not connected to the actual performance of the generator during shortage system conditions. On the contrary, the level of payment is determined by the regulator based on the physical characteristics of the generator. Therefore, there is no incentive for the generator to make efficient operational decisions to improve its availability during scarcity conditions and increase the reliability of the system.

2.4. Installed Capacity (ICAP) Markets:

The deficiencies of LMPs as signals for new investments have led to reliance on other mechanisms for resolving the problem of supply adequacy. FERC's SMD endorses the use of markets for installed or available capacity [ICAP/ACAP] based on requirements imposed by the ISO. In such a system, each LSE must provide evidence to the ISO that it owns or has contracts or credits for sufficient capacity to cover its peak load plus a reserve margin. LSEs and generators without sufficient contract cover can trade credits on generation capacity in a monthly market, thus enabling the LSEs to provide the required evidence to the ISO.

If LSEs fail to meet their ICAP obligations are subject to a deficiency charge (typically based on the cost of building a new capacity peaking generator). Generators used for providing ICAP credits must be physically capable of generating energy during a specified percentage of the year. ICAP generators typically must offer their energy for sale into the ISO's energy market and they are compensated for the reserve margins they provide. In economist's terms, the ICAP payment internalizes the externalities that are making the energy-only markets unstable and inefficient.

ICAP markets are notoriously volatile, since within the short time frame of a month the supply and demand are both highly inelastic. Another fundamental problem of the ICAP markets is their lack of linkage with the energy markets. ICAP markets and the trading of the reserve capacity requirement may produce prices that are not in equilibrium with the energy market prices. One way to resolve this problem is to require the ICAP contracts to have performance obligations requiring the capacity to be bid or to provide energy at some specified price. Therefore, if a generator receives a capacity payment, it must offer that capacity into the energy markets. Often, the requirement to offer capacity is independent of the price level at which energy is offered. This lack of a linkage has resulted in generators offering energy at extremely high prices in order to either exercise market power or to participate in other markets.

2.5. Available Capacity (ACAP) Markets:

Another way to tie value to payments is to base payments on generator unit availability as opposed to installed nameplate capacity. The California ISO has proposed an ACAP type of obligation in order to address some of the shortcomings of ICAP. The capacity product in this type of market is defined as the available capacity, which must be offered in the day-ahead energy market.

The ACAP product will be of various durations and can be provided either through generation capacity or physical load management. The ACAP obligations do not have a specified energy price ceiling, which currently is by default the regional price cap mandated by FERC. This price cap, however, is subject to change. Uncertainty in the price cap makes the pricing of long-term ACAP contracts difficult. Furthermore, since capacity is tied to specific assets, a secondary market for capacity is unlikely to arise, and if it does, there won't be any liquidity. Prices in such a market would not be very meaningful and the transaction costs would be extremely high.

Both ICAP and ACAP markets can suffer severely from the exercise of market power. The market price should reflect the marginal cost of providing capacity, which depends on the nature of the capacity obligation. Market power could be particularly acute if the obligation must be met near the date of delivery and available capacity is near the aggregate obligation level. One way to mitigate this problem is to require the capacity obligation to be met in advance of the delivery date rather than near the delivery date. This solution would also provide the market participants more options. The obligation could be met not only by existing generators but also by new entrants who could construct new capacity in time for delivery.

With the exception of very few capacity markets, such as in Colombia, it is striking that no other capacity market provides incentives for capacity transactions greater than a year.

3. OVERVIEW OF THE CAPACITY MARKETS IN THE US

The purpose of this section is to briefly compare the capacity market structure and rules of three Eastern ISOs: the PJM Interconnection ("PJM"), the NYISO, and the New England ISO. These systems transformed from tight power pools into markets. The origins of the current capacity markets evolved from historic utility practices of planning for adequate installed generation reserves.

3.1. PJM Interconnection

PJM has initiated a monthly or multi-monthly capacity market combined with a daily market that allows buyers and sellers of capacity to submit bids and offers. The daily market has been more liquid than the monthly market and prices in the daily auctions have been significant lower. Figures 1, 2 and 3 illustrate the ICAP quantities and prices from these markets.

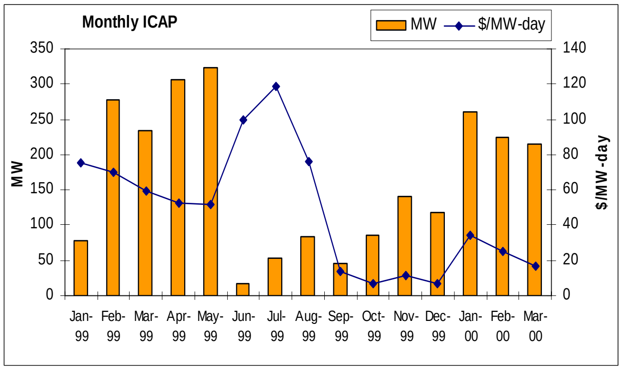

Figure 1: Monthly ICAP auction results

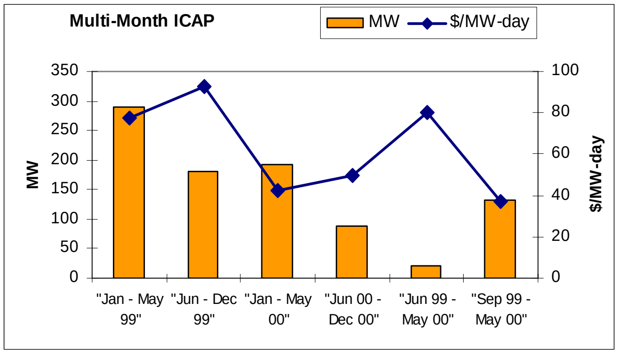

Figure 2: Multi-month ICAP auction results

Each LSE is required to install or contract for Capacity Resources or obtain Capacity Credits

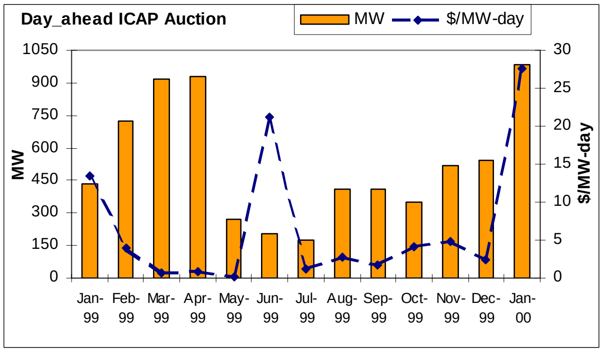

Figure 3: Monthly averages for the DA ICAP auction.

providing Unforced Capacity sufficient to satisfy each day its obligation. LSEs satisfy about 95 percent of their capacity obligations through self-supply and bilateral contracts, and the remaining 5 percent is traded on the Capacity Credit market. New LSEs rely more heavily on the market to satisfy their capacity obligations. If an LSE does not meet its ICAP obligation, it must pay a Capacity Deficiency Charge Rate (CDR) that is set currently at $58/KW-year. Revenues collected from the CDR charges are allocated to generators with long ICAP positions.

The PJM-ISO was the first power system to measure capacity on an unforced basis, thereby explicitly incorporating unit unavailability into the accounting of generation reserves. Over the last few years the PJM capacity market has undergone several modifications to address the following two problems a) market behavior, and b) unreliability due to capacity resources shifting between PJM and neighboring markets.

The original rules provided little incentive for LSEs to arrange for long-term capacity. In response to the perceived flaws in the capacity markets, PJM recommended certain tariff changes to the FERC related to a) adjusting the time period over which an LSE must commit generation resources to PJM, b) applying the deficiency charge on an interval basis, and c) requiring generation owners to commit excess capacity for an entire season, rather than on a daily basis.

3.2. New York Independent System Operator

The NYISO has an installed capacity reserve margin requirement for the New York Control Area based on the annual peak and a reserve margin. The NYISO has made two major changes from the original market design consistent with PJM's use of unforced capacity and facilitation of monthly capacity transactions.

The NY capacity market is split regionally, with New York City comprising its own reliability area. Starting at 2002 the NYISO moved from a 6 month to a one month obligation procurement period. However, the NYISO still holds a six-month "strip" auction for ICAP credits, and then monthly auctions to settle load shifts and deficiencies. The maximum deficiency rate is $13.67 per kilowatt-month in the New York City zone, $12.33 in Long Island and $10.50 elsewhere in the State.

3.3. New England ISO

On April 1, 1998, ISO-NE initiated an ICAP market. Each market participant is allocated a capacity requirement. There are two means for market participants to satisfy their responsibility: self-supply or bilateral transactions. ISO-NE also conducts a supply auction around the middle of each month to allow purchase and sale of UCAP for the following month. If, after the supply auction, ISO-NE determines that any LSE has failed to procure sufficient UCAP to cover its monthly requirement, it will conduct a deficiency auction. Participants are required to offer any UCAP that is in excess of their UCAP requirement in the deficiency auction. If a participant is still deficient after the deficiency auction, the participant must pay a deficiency charge. Most ICAP requirements are met through either self-supply or bilateral contracts. Small amounts are traded through the supply and deficiency auctions. Table 1 shows sample clearing prices for the ICAP auctions.

Table 1: ICAP Auction Clearing Prices

Obligation Month	Supply Auction Clearing Price ($/MW-Month)	Deficiency Auction Clearing Price
April 2003	$400.00	$0.00
May 2003	$150.00	$0.00
June 2003	$200.00	$0.00
July 2003	$200.00	$0.00
August 2003	$230.00	$0.00
Sept. 2003	$195.00	$0.00

4. PROPOSED METHOD FOR ENSURING ADEQUATE SUPPLY

The premise of the proposed method is that forward markets and hedging instruments can provide a market alternative to capacity payments. The design of ICAP/ACAP markets can be substantially improved by changing the product traded from capacity credits, which are merely paper "chits" showing that the capacity exists, to actual credits on energy output. These credits can take the form of fixed-price contracts, or preferably, call options or one-sided contracts-for-differences [CFDs]. In such a system the supplier reimburses the LSE if the market price exceeds the contract's strike price (see figure 4).

In order to assure that generators fulfilled their contracts, capacity payments would be made upon delivery of energy when called. However, a payment schedule needs to be developed for circumstances when the call option was not exercised. For instance, the payment schedule might require that not less than 75% of the annual payment be made by September 30 and 100% by December 31, regardless of how many times the option was exercised.

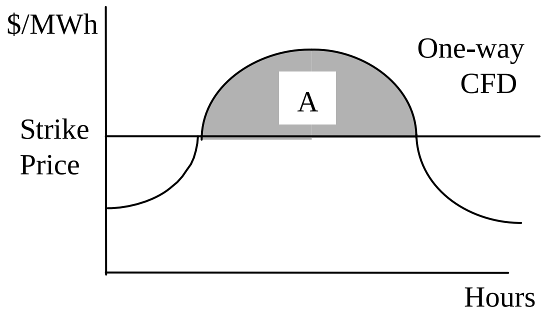

Figure 4: A: Seller pays buyer

From an engineering perspective it may seem sufficient simply to require each LSE to demonstrate the existence of adequate capacity to cover its peak load plus reserves. But the commercial implications of this policy are significant, and they affect the extent to which the policy succeeds in promoting its intended goal. An alternative policy focuses instead on energy availability rather than capacity. An LSE is required to provide evidence that it has long-term call options that enable it to call sufficient energy to meet its peak load and reserves. These options are usually written as one-sided CFDs, in which the LSE is hedged against market prices higher than the strike price. Actual call options would affect the energy market quite differently since the LSE would want to call the option whenever the market price exceeds the strike price.

This proposal has two advantages. One is that it addresses the actual need for energy in peak conditions, and moreover, at a "strike price" for called energy that reflects the long-term elasticity of supplies rather than the volatile short-term market prices in peak conditions. Once the options are exercised, one recognizes their second advantage. This is that the option portfolio's pattern of strike prices can be designed to induce price elasticity into the demand curve net of energy called via options, and thus reduce the incidence and magnitude of price spikes and mitigate the market power of generators.

With the proposed methodology, the capacity payments can be viewed as premium for call options corresponding to a relatively high strike price that will serve as a cap on the SMP. The strike price corresponding to the capacity payment should be higher than any marginal cost and be uniform for all generators for a total quantity that represents a target capacity for the system. The benefit of multiple strikes prices in mitigating price volatility also applies to contract duration. Longer duration contracts also mitigate price volatility and provide a stable income. On the other hand, short time horizons make the option premium fees to follow the pattern of the spot prices too closely. A one-year horizon is recommended for a general case, unless specific hydro cycles require larger periods. For systems with varying hydro years a three year time horizon is recommended.

In case a generator does not have the flexibility to shut down production due to minimum load conditions or due to take-or-pay fuel contracts a two-sided contract may be preferable. A two-sided CFD (i.e., a bundle of a one way call CFD and a one way put CFD) provides cover for producing at marginal cost that exceeds the SMP (see figure 5). Ideally it would be better to offer both one-sided and two-sided contracts so that parties can decide for themselves which contract better matches their risk management needs.

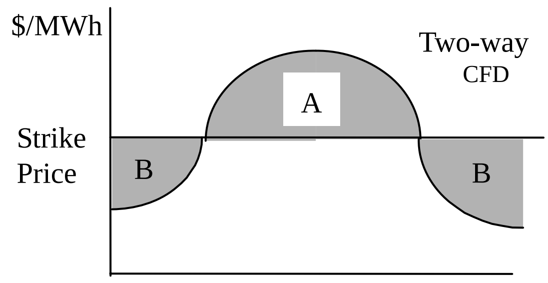

Figure 5: A: Seller pays buyer
B: Buyer pays seller

For the purpose of establishing capacity prices a single auction should be conducted for long-term call options. Strike prices can be a variable in the auction. Alternatively, the strike price can be determined by the regulator or the ISO. Other parameters that need to be determined in the auction are the volume of capacity to be auctioned off, the time horizon and the explicit penalty for non-compliance. The initial date of delivery also needs to be specified. This date must be far enough in the future to allow construction of new capacity, perhaps three years. The ISO could modify this time period as it gains experience with the market.

Furthermore, bidders would specify a delivery point into the system. Bidders who are looking to serve transmission constrained areas would use the delivery point to indicate their desire to serve a constrained area. Bidders who choose not to specify a delivery point would be obligated to arrange for delivery to a specified bus. This bus would be chosen by the ISO and would be designed to represent a reasonable trading hub.

Auctions that have the strike price specified by the regulator or the ISO may be sufficient to provide for capacity payments in order for the generators to cover their fixed costs. However these auctions may not be sufficient when an additional objective is to provide hedging against price volatility. Having the strike price predetermined, instead of allowing each generator to submit a strike price bid in the capacity auction, significantly simplifies the scoring rule of the auction. Some market designers argue that the value of the strike price is not critical since it only represents the demarcation line between the fraction of the generator's income that is recovered from the spot market and the fraction that is recovered from the call option premium. If the strike price is set administratively, it should be set high enough to ensure that the capacity insurance mechanisms associated with the call options only activate when the price is really high and, therefore, it does not interfere with the spot market under normal conditions. A value of the strike price that is at least 20% above the variable cost of the most expensive generator expected to produce is recommended.

The problem with the fixed strike price auctions is how to compare bids for fixed payment across generators with different marginal costs and how to induce generators to reveal their marginal costs. While it is possible to standardize many of the auction attributes (e.g., contract length, etc.), strike price and fixed payment are fundamental attributes that differentiate technologies and generators. Therefore a variety of options may need to be accommodated to provide flexibility to the market participants to be able to express through their bids their risk preferences. Ideally, it is desirable to allow generators to bid a strike price and a fixed payment and be able to select the most efficient portfolio of contracts. The key issue is to determine a scoring rule that will combine the two bidding dimensions.

5. CONCLUSIONS

This paper describes the various options available for ensuring adequate generation supply. It argues that energy-only markets are not sufficient to provide correct price signals to ensure adequate generation supply. Furthermore, they can be unstable and adversely impact the reliability of the system. It briefly presents the capacity markets in PJM, NYISO and NE-ISO. Finally, it presents a market-based approach that can play the role of ensuring adequacy. Under this method, an LSE is required to provide evidence that it has long-term call options that enable it to call sufficient energy to meet its peak load and reserves. These options are usually written as one-sided CFDs, in which the LSE is hedged against market prices higher than the strike price. The advantages and the various design options of the proposed methodology are described in detail in the paper.

FURTHER READING

New York Independent System Operator, 2002, *NYISO Installed Capacity Manual*, Schenectady, NY, August 20.

PJM Interconnection, LLC 2002a, *PJM Manual for Emergency Operations*, Manual-13, Norristown, PA, April 1.

US Federal Energy Regulatory Commission 2002a, ISO New England, Inc., *Order Accepting Amendments*, Docket Nos. ER02-1149-000, and –001, Washington, DC, July 31.

S. Oren and P.T. Spiller, "Wild prices out west: what can be done?", *Public Utilities Fortnightly*, November 1, 2000.

H. P. Chao and R. Wilson "Priority Service: pricing, investment, and market organization," *American Economic Review*, Vol. 77, No. 5, December 1987.

C. Vazquez, M. Rivier, and I. Perez-Arriaga, "A market approach to long-term security of supply," IEEE Transactions on Power Systems, Vol. 17, No. 2, pp. 349-357, May 2002.

BIOGRAPHY

Dr. Alex D. Papalexopoulos is president and founder of ECCO International, an Energy Consulting Company that provides consulting services on electricity market design and software issues within and outside the U.S. to a wide range of clients such as Regulators, Governments, Utilities, Independent System Operators, Power Exchanges, Marketers, Brokers and Software vendors. ECCO International is currently involved in various energy restructuring projects around the world including North America, Europe, Asia and South America. Dr. Papalexopoulos received the Electrical and Mechanical Engineering Diploma from the National Technical University of Athens, Greece and the M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, Atlanta, Georgia. He has published over 100 scientific papers in IEEE and other Journals and is the 1992 recipient of PG&E's Wall of Fame Award, and the 1996 recipient of IEEE's PES Prize Paper Award. Dr. Papalexopoulos is a fellow of IEEE.