Market Design and Pricing Incentives for the Development of Deregulated Real-Time Load Responsiveness Markets

*By Kenneth Skinner**

Introduction

Because of recent price volatility and resulting high prices, there has been a renewed interest in the consequences of supply and demand imbalance. The supply response is to build new generation. However, adding supply alone will not solve all of the problems, especially those associated with extreme price spikes. Both supply and demand responsiveness need to be addressed. On the demand side, market participants and independent system operators are reexamining the incentives and steps necessary to develop market-based demand responsiveness. In regulated markets, the cost and responsibility of Demand Side Management (DSM) programs were built into the rate-base or funded through green energy surcharges. In deregulated markets, where DSM programs or renewable energy investment must be recoverable through market-based pricing, these programs have been considered uneconomic and thus neglected.

In this paper I consider the necessary steps required of an effective and functioning real-time load curtailment market. Clearly, legislators and market participants need to re-focus on demand-side incentives. However, the issue is not so much whether these should exist, as how to create a competitive market where demand-side offerings are appropriately priced. First, in a deregulated market, the cost of demand-side programs must be recoverable through the offerings, not built into the rate-base. Second, market rules should be designed to allow free entry of competing suppliers of demand-side offerings. Third, care must be given to assure that the retailer bearing the cost is compensated, regardless of where the load reduction actually occurs. Finally, and perhaps most importantly, the current technique of load profiling must be redesigned to identify peak hour load reductions and compensate end-users appropriately.

In addition to market design issues, the paper further suggests a market-based method of pricing real-time load curtailment based on real-option valuation. The promise of real-time load reduction can be thought of as a strip of European call options. The strike-price is given by a contractually agreed upon threshold price between the energy provider and energy consumer. From price volatility determined from historic price data or implied from forward markets, a premium value is calculated for the right to curtail future load. Option premiums, profit sharing and limit orders can provide financial incentives for functioning demand responsiveness markets.

Supply and Demand Imbalance

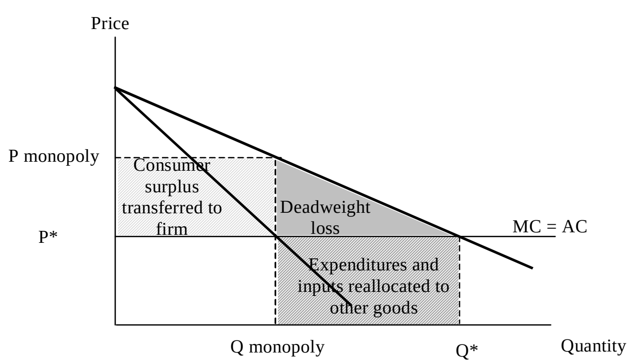
Because of recent price volatility occurring in deregulated wholesale power markets, legislators have begun questioning the fundamental reasons originally given for deregulating the electric utility industry. Early on, those favoring deregulation pointed to the advantage of perfectly competitive price determination in anticipation of lower energy costs.

However, in order for perfectly competitive prices to develop, fundamental assumptions of competitive markets must be met. One of these assumptions—the ease in which firms are able to enter markets—plays an important role in the development of competitive markets. Market entry assures that 1) long-run profits are eliminated by the new entrants as *prices* are driven to be equal to marginal cost, and that 2) firms will *produce* at the low points of their long-run average cost curves. Even in oligopolistic markets, long-run profits and prices exceeding marginal cost can be eliminated if entry is costless.

The recent California experience has highlighted the full extent of barriers facing new generation, and the cost to society when entry is constrained. In discussing the price setting power of monopolies, Nicholson (1992) states "The reason a monopoly exists is that other firms find it unprofitable or impossible to enter the market. Barriers to entry are therefore the source of all monopoly power" (p. 559). Figure 1 demonstrates the affect of market power in reducing output below optimal levels and raising market price to capture consumer surplus.

Because of decreasing economies of scale characteristic of large coal-fired steam facilities, electric utilities have traditionally been thought of as natural monopolies. If at any time due to transmission constraints, forced outages, or collusion amongst market participants (as was the case in the well-documented UK experience) a generator is able to command monopolistic power, prices will exceed marginal cost and consumer surplus will be transferred to monopoly profits.

Figure 1
Monopolistic Pricing



Only recently have electricity markets been contestable. A recent EIA (2000) report noted that with the exception of comparing variable operations and maintenance costs at nuclear plants to that of combined-cycle units, "the capital costs and both the fixed and variable operations and maintenance costs of combined-cycle plants, and conventional and advanced combustion turbines, are lower then the traditional baseload coal and nuclear technologies." (p. 42). As smaller units begin to compete with large baseload facilities, the market can no longer be characterized as a natural monopoly. Thus, significant advances in technological innovation have opened the door for competitive market pricing. H.R. Linden

* Kenneth Skinner is Manager of Derivatives Structuring, SEMPRA Energy Solutions, San Diego, CA. He can be reached at kskinner@semprasolutions.com This is an edited version of his paper presented at the 24th Annual IAEE Conference in Houston, TX, April 25-27, 2001.

(1997) noted in the *Electricity Journal*, "Under pressure of competition, the all-in cost of a combined-cycle plant has dropped to $450 per kilowatt, less than half that of a new clean coal plant. In combined-cycle configurations, heat rates have dropped. This has made natural gas at $2.50/million Btu competitive with coal in terms of variable cost when the much lower non-fuel operating and maintenance costs of gas are figured in."

However, until the barriers to entry are relaxed, prices will not be set at marginal cost. Because entrepreneurial merchant generation is unable to quickly enter the market to capture excess rents, existing generation is able to charge prices exceeding marginal cost.

There are several reasons why entry is constrained including site development and permitting delays, turbine availability and construction lead-time. Both advanced and conventional combined-cycle technologies require 3 years construction lead-time, while coal and nuclear plants require 4 years.[1] Once the facility is built, transmission rights and fuel availability constraints can limit market participation. Finally, scheduled maintenance and physically operating constraints can limit real-time market participation. It is apparent that physical generation by itself will not provide real-time market entry and exit required to assure marginal cost pricing.

Real-Time Load Responsiveness Market

In this paper, I suggest that the solution to costless entry is found in the "negawatt" market of real-time load curtailment. Unfortunately, effective programs designed to encourage active negawatt markets are only beginning to develop. A recent study by E SOURCE (2001) noted "As the electricity and gas industries struggle to take their first competitive steps, new pricing approaches will necessarily emerge, offering end users the opportunity, at least theoretically, to select the right product at the right price for them, as opposed to being subjected to the "class-average" tariff. But so far, research conducted by E SOURCE has uncovered few examples of pricing innovation in those regions that now have open access. In fact, regulated utilities may be more creative in providing options to their large end users—something quite unexpected given the flexibility open markets possess."

Theoretically, real-time load management is analogous to physical ancillary generation markets. Rather than dispatching and curtailing generation, real-time load management curtails and dispatches load. However, due to the high cost of monitoring and telemetry equipment and current limitations in market design, practical real-time load management is only available to large industrial consumers.

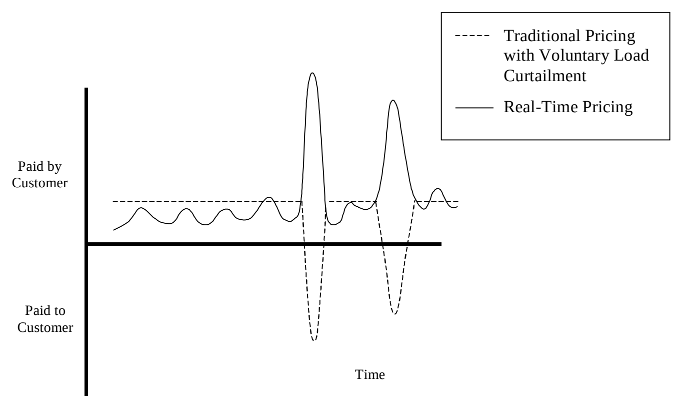
However, residential consumers can also participate in load curtailment markets. Residential customers can be encouraged to shift demand from peak to off-peak hours via a multi-tier tariff. For example, a simple two-tier system that prices peak power consumption differently from off-peak would provide incentives to shift non-essential activity to off-peak hours. Although limited, the opportunities for residential consumers provide a significant potential source of peak-load reduction. However, the current system of load profiling is fundamentally inconsistent with real-time load measurement and pricing. Until communities or entrepreneurial service providers commit to investing in multi-tier load monitoring, residential participation in load curtailment markets cannot develop.

Demand responsiveness markets will be most effective when shedding peak-load. E SOURCE (1999) demonstrated that small demand reductions could effectively bring wholesale prices way down. In many service territories, peak demand for the system, which may represent only 100 hours or so per year, creates the need for 10 to 25 percent greater system capacity.[2] In order for peak load shedding markets to develop, peak load price signals must be passed to end-use customers. As price signals become apparent, more end-users will find the flexibility and desire to sell back megawatts into the grid.

Current load curtailment programs are designed to benefit both the energy service provider (ESP) and the energy consumer. State regulators and ISO's encourage the programs. However, due to the cost of administrating the programs, the ESP must retain a large portion of the benefit in order to breakeven. Additionally, end-use customers tend to be risk-adverse when threatened with full exposure to real-time spot markets.

The most successful programs avoid much of the downside price risk through voluntary participation. Instead of threatening users with possibility of extreme energy costs, voluntary programs entice them with rewards for curtailing usage. These programs pass the price signals to the consumer, and, therefore, the incentive to curtail. However, if the consumer chooses not to respond and continues current consumption, they pay the conventional stable rate for electricity. Under voluntary load curtailment, shown in Figure 2, the energy user pays a standard rate that is designed to average out the highs and lows, but during a price spike event, the user can "sell back" the curtailed energy to the ESP. [3]

Figure 2
Voluntary Load Curtailment Pricing



As previously noted, current voluntary curtailment programs benefit both the ESP and the energy consumer though revenue sharing. The arrangement accounts for the shared risk and administrative expenses incurred by the ESP. However, other than for recovering administrative expenses, the ESP can be a neutral participant in the negawatt market. A functioning real-time negawatt market would automate much of the demand response activity. First, the energy consumer would determine ahead-of-time the strike price and

(continued on page 12)

[1] See footnotes at end of text.

Real-Time Load Responsiveness Markets *(continued from page 11)*

level of curtailment consistent with their opportunity costs. The strike price would then be compared to expected system price on a day-ahead and hour-ahead basis. If the expected system price exceeds the strike price, the customer is automatically notified. Ultimately, the real-time transition from system energy to backup-generation would also be automated. The negawatt market participant would automatically transition off of system load.

Competing with Generation Companies

Ideally, the ESP would be indifferent to either paying GenCo's the spot market price for wholesale energy or paying the negawatt participant for load curtailment. Under this scenario, the end-use customer receives the full benefit of equivalent spot market prices for participation in the negawatt market. The benefit to the ESP is less apparent. If the load curtailment generates enough savings, the market would face a less expensive marginal unit setting market price. In this case the ESP would receive a higher return on power sold to fixed tariff customers.

Load responsive negawatt markets can provide system capacity through either reducing consumption or switching to backup-generation. For the purpose of calculating the cost to shed system load, the two options are equivalent. Both switching to backup-generation and shedding load represent opportunity cost. However, the advantage of focusing on the cost of backup-generation is that it effectively sets an upward bound on cost. The annualized cost of backup-generation effectively caps the power market annualized price. At the point where system cost exceeds the cost of new generation, negawatt market participants would be better off installing new backup-generation than purchasing from the power market. Negawatt markets would compete directly with GenCos, creating a demand response cap to market price and volatility.

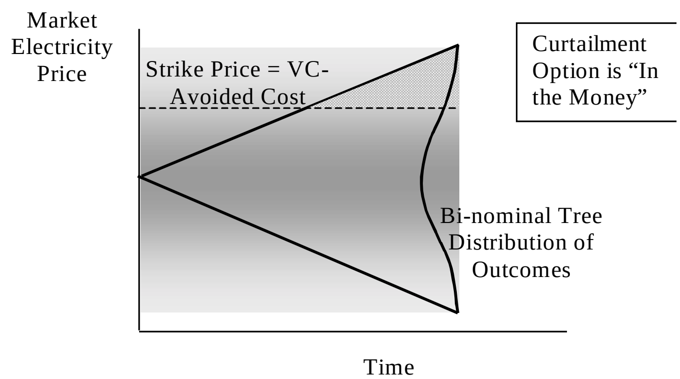
Although negawatt market participation can be either through reducing consumption or switching to backup-generation, for the purpose of market pricing, we consider all participation as if though backup-generation.

Real-Option Pricing

Using real-option valuation of participant opportunity costs, price incentives exist for negawatt market development. The opportunity to switch from system load to backup-generation may be modeled as a series ("strip") of options on Btu spreads, and option valuation techniques employed. Figure 3 represents the possible outcomes of valuing a negawatt participant strike price for real-time market entry. The figure demonstrates how the end-user determines at what point to sell back to the negawatt market. That point is the strike price at which the end-user exercises the option to participate in the negawatt market. The strike price is the variable cost of backup-generation less system power purchase costs. At the strike price, the participant is better off running backup-generation and collecting market revenue for its equivalent capacity contribution, than purchasing energy from the retail energy market.

The option value is equivalent to the amount an end-user would be willing to pay in order to participate in the negawatt market—the net cost of backup-generation. The approach effectively caps system volatility and peak-price. Volatility and peak price would never exceed the cost of installing new backup-generation (less avoided system cost).

Figure 3
Pricing the Real-Time Market Entry "Strike Price"



The Btu spread associated with Figure 3 represents the differential between electricity and fuel prices, in Btu-equivalent measures. Such a spread is most commonly calculated between electricity and natural gas, and known as the spark spread. In our case, we are considering the spread between backup-generation fuel oil and electricity. Prices are adjusted for heat rate. Thus, each curtailment market backup-generating unit has its own spark spread. The spread is location-specific, and the adjustment factors may possibly take into account location-specific, transportation over pipelines and electricity transmission lines.

The spark spread may be positive or negative. When the spark spread is positive, it means that fuel oil is more valuable burned for electricity by backup-generation than as a raw commodity. When the spark spread is negative, it means that the fuel oil a generating unit burns is more valuable than the electricity the unit produces. An arbitrageur would pay an end-use customer with a long-term fuel contract not to operate in such cases, but to give its fuel over to the arbitrageur for sale in the commodities market. In essence, when a generating unit's spark spread is negative, its generating capacity has no immediate value in the energy market.

An electric generating unit can be thought of as a means to capitalize on the spark spread. When the unit's spark spread is negative, the curtailment market participant should purchase its power from the retail energy market. When the unit's spark spread is positive, the market participant should operate its backup-generator in direct competition to the power generation companies.

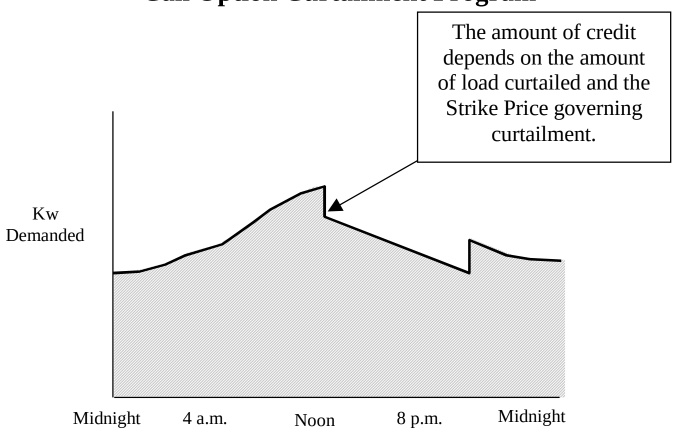
However, to burn fuel oil for electricity requires having backup-generating capability available. While an arbitrageur trying to take advantage of a negative spark spread need only to find a buyer (and associated transportation) for the fuel, to take advantage of a positive spark spread an arbitrageur needs backup-generating capacity (or the equivalent ability to reduce power consumption). If such backup generating capacity were instantly available and costless, then arbitrage would drive a positive spark spread to zero effectively capping energy market prices via market participation.

The Cinergy Baseline Reduction Program

Although competitive negawatt markets do not currently exist, entrepreneurial energy service providers are currently using a real-options theory to value curtailment products. The Cinergy Baseline Reduction Program is one example. Participants in this program are able to choose the level of risk that curtailment will occur and the amount of energy curtailed. Choosing a lower Strike Price increases the possibility of curtailment.

Participants receive a corresponding premium payment and an energy credit for curtailed energy. The premium payment is based on the Strike Price, the option load contracted, and the operational plan selected. A "Call-Option" in this case gives the ESP the right to purchase energy from the end-use customer at the agreed upon Strike Price. The Call Option is exercised when the ESP marginal cost of electric energy, including all variable cost associated with delivering the energy, is projected to be equal to or greater than the Strike Price. Figure 4 represents how end-user load shape responds to the Call Option.

Figure 4
Call Option Curtailment Program



The Cinergy load curtailment program contains many of the elements necessary for negawatt market development including option pricing, risk sharing, voluntary participation, ESP customer support, and reliability. Such programs will provide the foundation for development of real-time load responsiveness markets.

Conclusion

Theoretically, real-time load management is analogous to physical ancillary generation markets. Rather than dispatching and curtailing generation, real-time load management curtails and dispatches load. Responsive load "negawatt" markets can be developed to create real-time entry and exit fundamental to competitive priced electric power markets. Negawatt markets would compete directly with GenCos, creating a demand response cap to market price and volatility. Generators would compete with backup-generation, the cost of which sets the market cap.

Using a market-based method of pricing real-time load curtailment, based on real-option valuation of participant opportunity costs, price incentives exist for negawatt market development. The promise of real-time load reduction can be thought of as a strip of European call options. The strike-price is given by a contractually agreed upon threshold price between the energy provider and energy consumer. From price volatility determined from historic price data or implied from forward markets, a premium value is calculated for the right to curtail future load. Option premiums, profit sharing and limit orders can provide financial incentives for functioning demand responsiveness markets.

Footnotes

[1] Energy Information Administration, *Assumptions to the Annual Energy Outlook,* DOE/EIA-0554 (Washington DC, January 2000), Table 37, Cost and Performance Characteristics of New Central Station Electricity Generation Technologies.

[2] David Stern, Manager, Product Support, POWERdat/BaesCase, Resource Data International, FT Energy, 3333 Walnut Street, Boulder, CO 80301, tel 720-548-5427.

[3] Dick Montague, "Voluntary Load Curtailment Systems for Win-Win Load Control," E SOURCE Report EIC-13 (December 2000).

References

Capage, A., R. Davis, and W. LeBlanc, 1999. "The Dawning of Market-Based Laod Management," E SOURCE *Report ER-99-18* (November).

Energy Information Administration, *Assumptions to the Annual Energy Outlook,* DOE/EIA-0554 (Washington DC, January 2000), Table 37, Cost and Performance Characteristics of New Central Station Electricity Generation Technologies.

Energy Information Administration, *The Changing Structure of the Electric Power Industry 2000: An Update,* DOE/EIA-0562(00) (Washington DC, October 2000).

Hensler T., and W. LeBlanc, 2000. "Energy Pricing and Load Management: What Do End Users Want?" E SOURCE *Multi-Client Market Research Study* (March).

Linden, H.R. 1997. "Operational, Technological, and Economic Drivers for Convergence of the Electric Power and Gas Industries." *The Electricity Journal* (May).

Montague, D. 2000. "Voluntary Load Curtailment Systems for Win-Win Load Control," E SOURCE Report EIC-13 (December).

Nicholson, W. 1992. "Microeconomic Theory: Basic Principles and Extensions. Dryden Press, Orlando, FL, p. 556.

Aberdeen Program *(continued from page 5)*

with KLM/Northwest for conference delegates. For full details see the special website at www.abdn.ac.uk/iaee. The airport is 20 minutes drive time to the City Centre or the Conference Centre. There are direct train links from London and many other cities in the UK to Aberdeen.

Queries: (excluding Social Tours)	**Social Tours:**
Professor Alex Kemp	Pam Wells
Department of Economics	Corporate Events
University of Aberdeen	Gowanbank
Edward Wright Building	Station Road South
Dunbar Street, Old Aberdeen	Peterculter
AB24 3QY Scotland, UK	Aberdeen
Phone: +44 (0) 1224-272168	AB14 0LL Scotland, UK
Fax: +44 (0) 1224-272181	Phn/Fax:+44(0)1224-735733
Email: a.g.kemp@abdn.ac.uk	wellspj@compuserve.com

**** CONFERENCE SPONSORS TO-DATE**: Shell, BP, UK Department of Trade and Industry, Aberdeen City Council.