The Use of Marginal Energy Costs in the Design of U.S. Capacity Markets

Robert Moye Sean Meyn*

Abstract

This paper surveys the development of marginal cost theories used in the optimal allocation of scarce resources, and examines the application of these theories to current-day electricity capacity markets. The different approaches in use today to ensure grid reliability and incentivize new resources are examined. Market challenges are surveyed, as well as empirical findings that suggest that current market approaches do not provide proper incentives. We conclude that the so-called "missing money" is not missing because of defects in market designs, or so-called administrative actions—money to incentivize investments is missing due to a misapplication of marginal cost theory.

Acknowledgements Research supported by the National Science Foundation under grants CPS-0931416 and EPCN-1609131. We thank Professors David Spence, Amy Stein, and Seth Blumsack for their insightful comments and encouragement. Their input as well as suggestions from the reviewers were carefully considered in the revised manuscript.

1 Introduction

> *Much of the social history of the Western world over the past three decades has involved replacing what worked with what sounded good.* – Thomas Sowell

Driven by claims during the last quarter of the 20th century of anticompetitive behavior by electric utilities, frustrations by consumers having to bear much of the risk of large electric generation investments, and a desire by many to create a more economically efficient market for the electric power industry, the United States began to deregulate (or "reregulate") certain aspects of the electric power industry. Starting with the *Public Utilities Regulatory Policies Act ("PURPA")* in 1978 and continuing with energy policy acts in 1992 and 2005, the U.S. wholesale power markets (and in some states, the retail markets) were opened to more competition. Much like the deregulation of AT&T in the 1980s, these actions were expected to spur innovation and reduce costs to consumers. And like the experience with AT&T, opinions on the results of the deregulation of the electric power industry are mixed.

Wholesale power markets in the U.S. today represent a patchwork of policies and approaches that impact, if not outright dictate, how capacity, energy and related products are bought and sold. While some aspects of all markets appear to function well, there continues to be much debate about the efficacy of some. This is particularly true for the markets designed to ensure that sufficient electric generation is in place today and being planned and constructed for future use. These *Capacity Markets* currently take many and varied forms across the U.S. Most markets, but not all, define a minimum *resource requirement* that entities directly serving customer load are required to meet. Some rely on bilateral markets to meet these requirements. Others have very structured processes (e.g., auctions) to facilitate the purchase and sale of generating capacity. And while there are many prominent individuals involved in the market-structure debate today who believe long-term markets will self-optimize if only we can be patient, this issue is far from settled.

There is little scientific basis to predict that long-term optimality will emerge from short-term decision making by generation operators. One challenge to analysis is that there is no agreement on how to quantify risk to society or to an individual agent in the market. Another challenge is the enormous uncertainty over planning horizons of many decades. The risk to a generator operator is obvious in today's technological environment: the lifetime of an efficient combined-cycle generating station may be a half century, and its purchase price over one billion dollars. At the same time, revenue over this period depends on uncertain energy prices and policy.

The remainder of this paper is organized as follows: The evolution of regulation of the electric power industry is discussed in Section 2 to provide a backdrop for the state of the markets today. Section 3 contains a short history of marginal cost theory and its use in today's energy and capacity markets. This theory is based on the notion of *efficiency*, whose definition is based on a hypothetical Social Planner's Problem.

*The authors are with the Department of Electrical and Computer Engineering at the University of Florida. RM is a doctoral candidate at UF and executive at Rainbow Energy rmoye@ufl.edu. SM is the Robert C. Pittman Eminent Scholar Chair at UF meyn@ece.ufl.edu.

URI: http://hdl.handle.net/10125/50214
ISBN: 978-0-9981331-1-9

Current market structures are surveyed in Section 4, with emphasis on the elements of mechanisms used to incentivize investment in generating resources. It is here where we find potential gaps between the hoped-for optimal Social Planner's solution, and the outcomes of markets in a real-world setting. Some of these shortcomings are discussed in Section 5, and potential solutions are presented in Section 6.

2 Evolution of the Power Industry

Today, electricity is so basic to the world economy that certain electricity indices are used to express a country's economic standing (consumption or production of electricity per capita) and the standard of living enjoyed by consumers (per capita electricity consumption in the domestic sector [16]. As such, the availability and cost of electricity is fundamental to the economic wellbeing and prosperity of a society.

Primarily as a result of competitive market forces, the electric power industry has evolved significantly over time. Generating resources have become more reliable and efficient. High-voltage transmission networks, nonexistent at the birth of the industry, are now extremely reliable and efficient.

Practices and procedures, both for system operation and for long-term planning, have also improved greatly and now contribute to the overall value and efficiency of the industry. While many improvements were realized in the early days through trial and error, today's systems benefit from the extensive use of computers to optimize both short-term operation and longer-term system expansion. These tools have been particularly helpful in enhancing short- and long-term planning techniques and practices.

The regulatory paradigm has also changed significantly over the history of the electric power industry—starting first with regulation by municipalities through the granting of franchises. This was followed by the creation of public service commissions in each state, and eventually regulation of wholesale market activities at the federal level. In general, these changes were made to protect electricity consumers from anti-competitive behavior [16].

In the second half of the 20th century, the industry was again changed to promote more competition. The Public Utilities Regulatory Policy Act of 1978 for the first time allowed companies other than regulated utilities to sell electricity in the wholesale power market (limited to renewable energy and cogeneration resources).

The passage of the Energy Policy Act of 1992 marked a significant evolution of the industry. Following development of rules by the Federal Energy Regulatory Commission (FERC), the high-voltage transmission systems that interconnect the utilities in the United States began providing open access to all existing utilities and wholesale generators, and non-utilities were allowed to own and operate electric generation for sale into the wholesale market. In addition, entities called power marketers (at the beginning of the electricity markets, typically affiliates of utilities and investment banks) could participate freely in the market by purchasing electricity from one entity and selling to another.

The re-regulation of the electric power industry led to regional organizations designated to operate the high-voltage transmission systems on a statewide or multi-state basis, and to implement electricity markets for the purchase and sale of electricity products. Underlying each of the markets is a particular mathematical formulation of efficiency, and surrounding marginal cost analysis is the core of this philosophy. The following section examines the foundation for these markets and the problems they create (see Section 4 for a discussion of these new markets).

3 Marginal Cost and Efficiency

> *[E]very tub must stand on its own bottom, and that therefore the products of every industry must be sold at prices so high as to cover not only marginal costs but also all the fixed costs, including interest on irrevocable and often hypothetical investments...* Hotelling [19, pg. 242].

Short- and long-term optimization of resources in today's *Organized Markets* lean heavily on marginal cost theory and the concepts of economic efficiency. A major weakness is their reliance on short-run marginal costs to provide long-run investment signals. A review of the research regarding the use of marginal costs to set the price for factors of production reveals that some other means of addressing the fixed cost of assets is needed; this fact was recognized by commonly cited authors in this field, such as Coase and Schweppe.

3.1 Marginal Costs

The discussion of the use of marginal cost pricing for public utility projects began with a French engineer in the 1800s. Jules Dupuit introduced the concept of marginal utility in an 1844 article concerned with the optimum toll for a bridge [12]. This theory was further formalized by Alfred Marshall in 1890 when he combined the ideas of supply and demand, marginal utility and costs of production [25].

In 1937, Harold Hotelling presented an update to the work of Dupuit (and used the supply and demand curves of Marshall) to argue, among other things, that the use of tolls on bridges in New Jersey was resulting in less-than-optimal use [19]. Hotelling argued that because the amount of the toll was above the marginal cost to allow people to use the bridge (which was essentially $0), it prevented some from utilizing the bridge that would otherwise benefit from such use (because their marginal value was above $0, but less than the amount of the toll).

In 1946, R. H. Coase addressed the issues presented by Hotelling and others and specifically focused on the "conditions of decreasing costs" [19, 20]; see also [24, 27]. Coase agreed that the amount paid for goods and services should equal the marginal cost to produce or provide the goods and services. However, he pointed out that whenever marginal costs are less than average costs, the total amount paid for a product will fall short of total costs. The is particularly relevant to the optimization of power systems, where average *total costs* are well above average marginal costs.

Marginal analysis was first applied to investments in electric power supply by the Electricité de France (EDF) in the late 1940s and in the 1950s. While most efforts in the United States were focused on the theoretical aspects of marginal pricing, EDF was concerned more with the practical implementation [30, 38]. This work led EDF to implement a transmission tariff in 1957 that utilized marginal cost pricing and incorporated these same concepts into long-term investments. Marcel Boiteux (during this time an engineer at EDF and later its Chairman), studied the relation between short- and long-run marginal cost pricing. The solution provided by Boiteux, et al, was to increase the price beyond marginal costs. EDF continues to lead in both the economics and the engineering foundations for long-term planning and investment [1]. Quantifying risk aversion and uncertainty is an essential component of this research.

In the 1970s, work in this area continued by Baumol and Bradford [3] and Feldstein [13], where Ramsey-Boiteux pricing[1] was used to derive how prices should be increased above marginal cost in order to meet "social revenue requirements." In 1971, Vickrey introduced the concepts of "real-time pricing" for a product, albeit for telephone service pricing [3]. However, it wasn't until the 1980s when work by Schweppe, et al, focused specifically on electricity [6]. This work, along with other work done by his co-authors [4, 5, 36], led up to the book that many today point to as the basis for the use of marginal cost pricing in Organized Markets - *Spot Pricing of Electricity* [34].

It is a crucial fact that all of the prior research and analysis into the use of marginal costs from Dupuit to Schweppe, et al, are consistent with the idea that while prices for electricity at marginal cost optimize the general welfare in the short-run, basing revenues entirely on short-run marginal costs is not sufficient to recover fixed costs, and therefore insufficient to incentivize investment in generation.

3.2 Social Planner's Problem on Engineering Timescales

Economic systems are said to be Pareto optimal if there is no alternative way to "organize the production and distribution of goods that makes some consumer better off without making some other consumer worse off" [26].

From a power supply perspective, a power system is said to be operating under optimal conditions if there is no alternative way to lower short-run[2] costs by redispatching or modifying the commitment of available generating resources. However, over the long-run a system can be said to be optimal only if investment decisions are also incorporated into the analysis. That is, a long-term power supply plan can be said to realize Pareto optimality only if there is no other combination of existing and potential resources, along with the optimal commitment and dispatch once given these resources, over the useful life of the resources. How can an approach that only uses short-run marginal costs, and ignores long-term investment costs, provide for an optimum system?

The primary challenge with incentivizing investments in today's electricity markets centers around the time frame covered by our decisions. Operating decisions are short term; from a few minutes to a few years. Investment decisions are long term; from a few years to several decades. More specifically:

- Decisions on the optimal use of generating resources to serve the expected load 5 minutes in the future consider only those resources currently on line and synchronized to the system.
- Decisions on the optimal use of generating resources to serve the expected load 30 minutes in the future consider those resources currently on line and synchronized to the system, plus those that can be brought on line and ramped up to provide the desired output within that

[1] Ramsey-Boiteux pricing is a policy concerning what price a monopolist should set, to maximize social welfare, subject to a constraint on profit.

[2] "Short-run" in this context refers to the period from the next five minutes through the next few years (i.e., as limited by the time it takes to install additional generating resources).

time frame.

- Decisions on the optimal use of generating resources to serve the expected load 3 days in the future consider almost all existing and available resources.
- Decisions on the optimal use of generating resources to serve the expected load 10 years in the future consider all existing resources, plus any resources and technologies that can be installed prior to this time. Also, resource overhauls, repowerings and retirements are considered over these longer time scales.

While there is theory to support the emergence of efficiency as the result of short-term optimization by selfish agents in the market, this theory is not likely to be predictive on the timescales of interest in this paper. ***We believe that a long-term investment scenario that is consistent with Pareto optimality can be achieved only with a certain level of long-term planning.***

4 Organized Markets in the U.S.

The electricity markets in the U.S. today can be viewed as falling into one of two paradigms. There continue to be "bilateral markets" in which buyers and sellers negotiate the purchase and sale of energy and capacity directly with each other.[3] These transactions can range in timescale from the next hour up to several decades, and the characteristics (e.g., firmness, delivery location and price) can be different for every transaction. And while under current regulations, any entity can participate in these transactions, it takes a certain set of knowledge and skills to be effective in this market.

Outside of the bilateral markets, and covering most of the U.S., Organized Markets have been established to provide for the buying and selling of energy, ancillary services and, in some cases capacity, via a central clearing mechanism. The primary purpose of these markets is to separate generation and retail electric service from the natural monopoly functions of transmission and distribution.

The primary agents in these models are generation companies that supply the electric power, and the *Load-Serving Entities* (LSEs) that are responsible for providing electric service to retail customers [14]. Examples include investor-owned electric utilities such as Pacific Gas & Electric, and not-for-profit community-choice aggregators (CCAs) such as Marin Clean Energy. The term "customer" is reserved for the end-consumer of electricity—residential, commercial or industrial.

4.1 Marginal Pricing in RTOs

Both Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) are organizations formed with the approval of the FERC to coordinate, control and monitor the use of the electric transmission system by utilities, generators and marketers. More specifically, an ISO, as specified in FERC Order 888, is a non-profit organization that is designed to provide non-discriminatory service to all market participants, and is independent of the transmission owners and the customers who use its system. RTOs, defined in FERC Order 2000,[4] also provide non-discriminatory access to the transmission network, but have some additional responsibilities dealing with transmission planning and expansion for the entire region served by the RTO.

Today there are nine ISOs/RTOs operating in North America. They manage the systems that serve two thirds of the customers in the U.S., and over half the population of Canada. Over time, the distinction between ISOs and RTOs in the United States has become insignificant. Both organizations provide similar transmission services under a single tariff at a single rate, and they operate energy markets within their footprints. For brevity, we refer to either ISOs or RTOs, or collectively Organized Markets, simply as "RTOs."

The *Locational Marginal Price* (LMP) is intended to be the cost of supplying, at least-cost, the next increment of electric demand at a specific location (node) on the electric power network, taking into account both supply (generation/import) offers and demand (load/export) bids and the physical aspects of the transmission system, including transmission and other operational constraints [37]. By design, when the lowest-priced electricity can be delivered to all locations in the market footprint (i.e., there are no transmission constraints), and ignoring electrical losses, prices are the same across the entire RTO. However, when power flowing over the transmission system reaches limits designed to ensure reliable operation, the lowest-priced energy cannot flow freely to some locations and more expensive generation is required to serve the load in the constrained regions. Under this scenario, LMPs are subsequently higher in those locations.

A key element of the structure of energy markets

[3] Bilateral markets for all capacity and energy products continue to operate in the southeast and parts of the western U.S. Most of the country continues to utilize bilateral markets for capacity—at least for meeting part of the markets' needs.

[4] While the functions of RTOs are similar to those of ISOs, FERC chose to use a new name in Order 2000 for its desired form of transmission organizations in the U.S.

within all RTOs is that resource owners and LSEs submit offers to sell and bids to buy hourly blocks of energy for all 24 hours of the next operating day. The RTO takes these offers and bids and determines the least cost, security constrained commitment and dispatch of resources to serve the LSEs for the next operating day. Out of this process, *day-ahead* LMPs are created from the prices offered and bid by the participants.

In addition to this *day-ahead market* is a *real-time market* in which LMPs are calculated every five minutes and represent the price that LSEs will pay and generators will be paid for the subsequent five-minute period. Even when subject to transmission limitations or ramping constraints on generation, prices in excess of marginal cost, and even *negative prices* are consistent with economic efficiency [11, 39].

4.2 "Administrative Actions"

Many entities involved in the RTOs (RTO staff, market participants, market monitors, regulators, and market advisors) believe that an economically efficient market is one in which the only compensation paid to a generator is tied to a markets' short-term LMP [8, 18]. Whether or not this is true, they universally recognize that problems with the markets' design keep the markets from operating efficiently. They believe these problems include:

1. A lack of direct participation by the customers within the same timescales as the generators (e.g., hourly).
2. The use of price caps to limit the maximum price an LMP can rise to, and thus limit the potential revenue a generator can receive.
3. The use of "administrative actions " by system operators to ensure reliable system operations. While by design, LMPs are not subject to manipulation by market participants, in practice system operators have substantial discretion over LMP results through the ability to classify units as running in "out-of-merit dispatch."[5] When this occurs, these units are excluded from the LMP calculation which often results in depressing market prices.[6]

The *missing money problem* often cited in capacity market research refers to a class of failures in organized markets. That is, expected net revenues from sales of energy and ancillary services are believed to provide inadequate incentives for investors in new generating capacity (or equivalent demand-side resources) to invest in sufficient new capacity to match administrative reliability criteria [22] *because* of these market failures. "The fundamental source of the net revenue gap problem is the failure of spot energy and operating reserve markets to perform in practice the way they are supposed to perform in theory [21]." The consequence is that prices paid to generators in the energy and ancillary service markets are substantially below the levels required to stimulate new entry.

Organized Markets have therefore been useful in bringing efficiencies to short-term system operations and dispatch, but, in the opinion of some [9], have been a failure in what was advertised as a principal benefit: stimulating suitable new investment where it is needed, and when it is needed.

4.3 Scarcity Pricing

Scarcity occurs when available generation is insufficient to cover the expected energy *and* operating reserves required for reliable operation. *Scarcity Pricing* provides for an increase in the LMP during defined scarcity conditions—such conditions being tied to the level of reserves (regulating, spinning, standby, etc.) available to be called upon if needed. As touted by the supporters of this approach, this is a means to stimulate a more competitive market and to better provide incentives for investments in supply-side and demand-side resources.

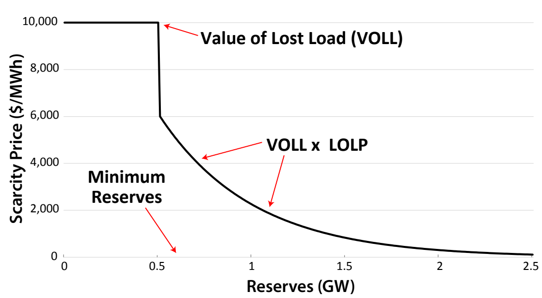

Figure 1: Operating Reserve Demand Curve

Some RTOs have implemented versions of Scarcity Pricing with the design of pricing mechanisms based on two concepts from traditional system planning: 1. the *Value of Lost Load* (VOLL), in units of \$/MWh, that is intended to represent the cost to the ultimate electricity consumers when load is interrupted, and 2. *Loss of Load Probability* (LOLP), defined as the probability that the entire load cannot be served.

As an example of one Scarcity Pricing design, ERCOT utilizes an *Operating Reserve Demand Curve* (ORDC) which adds a Scarcity Price to the

[5]This indicates that one or more resources being dispatched are done so for reasons other than economics.

[6]Under these circumstances, more expensive generation is brought on line but not allowed to set the market LMP. Other generators are then required to reduce output, further lowering the overall market LMP.

LMP during any defined periods of scarcity. Figure 1 illustrates the basic structure of the ORDC used in the ERCOT market [18].

The primary components of an ORDC include: i) a price, assumed equal to the VOLL, to be paid to all resources participating in the real-time market when operating reserves fall below a set level (assumed equal to the market's minimum operating reserve level); and ii) a price to be paid to all resources participating in the real-time market as operating reserves approach the minimum designated level.

Though the ERCOT report is not clear, it is assumed that the market operator would determine a real-time LOLP that corresponds to the settlement period in question. The VOLL is assumed to be applicable to all customer classes, and independent of time. Research has indicated a significant range for VOLL in practice, with values depending on the time of day, day of week, customer class and also on duration [23, 33]. Consequently, the real-time LOLP and stationary VOLL are parameters that the markets are currently not equipped to precisely define or calculate.

4.4 Incentivizing Investments

The methods used in the RTOs in the United States to incentivize investment in new generation fall into three categories:

- **Energy-only:** An approach wherein revenues from the energy markets (and ancillary services markets) are expected to provide sufficient compensation and price signals to optimize resource investments.
- **Energy + Capacity Markets:** The above energy-only pricing approach, plus formal capacity markets, which together are expected to provide sufficient compensation and price signals to optimize resource investments.
- **Traditional:** Least-cost, long-term resource planning methods that have been used by utilities for decades as "incentives" for investment decisions. For utilities whose generation investments are still regulated by state utility commissions, this approach involves demonstrating that a proposed expansion plan is "best" (e.g., least cost, or close thereto) of the plans evaluated.

Outside of the RTOs, the Traditional method is being employed, and with good success. Over the past 15 years, for example, the largest investor-owned utility in the state of Florida (Florida Power & Light, or FPL) has invested in some of the most efficient and cost-effective combined-cycle facilities in the country [15].[7] And while FPL still operates under a cost-of-service, rate-of-return paradigm, even with this significant expansion of its generation base, it has seen very little increase in retail rates over this same period.

5 Critique of Methods used to Incentivize Resource Investments

The fundamental assumption behind today's RTOs is that if competitive market forces are allowed to take place in the daily and hourly energy and ancillary services markets, the resulting prices (LMPs) will provide all of the incentives needed for supply side resources. That is, these markets alone can optimize the Social Welfare. The theory supporting this assumption is flawed, and moreover there is no empirical evidence that short-term markets will lead to long-term optimality.

5.1 Critique of Energy-Only Market Designs

For the energy-only market theory to work, fundamental changes need to occur in the market design.

1. **Eliminate Reserve Margin Criteria** – In [11, 39] it is shown that a competitive equilibrium will result in optimized reserves from the perspective of the social planner, and average prices will equal average marginal costs. This conclusion is valid even with the introduction of ramping constraints, and uncertain demand and supply. However, hiding behind these mathematical conclusions is this fact: there will be no long-term equilibrium since the generation companies will not be able to cover their fixed costs.

   It has been argued that the reserve margin criteria in place in the markets, wherein defined levels of generating capacity in excess of what is necessary to serve the forecasted peak load, must be relaxed, if not eliminated. Without this change, the markets will rarely be without sufficient installed capacity to serve peak loads.[8]
2. **Provide for Direct Retail Participation** – Retail consumers of electricity must be given the ability, and have the desire, to participate directly in the market. It is claimed that they must in fact respond in real-time to market price signals (see Borenstein's survey [7] and Wolak's testimony to the California state government [41]). While significant technological

[7]Since 2002, FPL has added over 15,000 MW of highly efficient, natural gas-fired generation.

[8]Current reserve margin requirements in place across most of the U.S. (12% to 20% of projected annual peak load) ensures that generation is available to serve load 99.97% of the time.

advances have been made (e.g., with *smart meters*), it is still difficult to see how price signals can be of value to customers, or of value to the grid operator [28, 29]; in particular, price signals cannot produce high quality balancing reserves or ramping services that can be obtained through distributed control [10].

3. **Eliminate Administrative Actions** – Administrative actions (market price caps and reliability-based, out-of-merit dispatches of resources) would have to be eliminated.

In the unlikely event that *any* of these changes could be made, we would still be challenged by other issues. Even if regulators go along with the plan, will customers accept the reliability construct required to create the price spikes needed to incentivize generation? Will the trigger prices set by retail customers for limiting service, equal or exceed those required by resources to be adequately incentivized? Will demand-side solutions crush any price spikes expected as a result of lower reliability standards? ***Like the Scarcity Pricing construct, without crisis there will be no opportunity!***

Because an energy-only structure would likely operate in a manner similar to that of today's Scarcity Pricing approaches, it is worthwhile to point out some significant short-comings in that pricing scheme.

1. Scarcity pricing in ERCOT and other markets is linked directly to conditions related to a lack of *operating reserves*[9] and not *planning reserves.*[10] Therefore, scarcity conditions occur when generation or transmission resources become unavailable or limited during the *operating conditions.* This rarely relates to the total capacity of resources installed in the market.
2. In the current market approach, Scarcity Pricing is only applied to generation and loads that deviate from the amount cleared in the day-ahead market. Therefore, any resources that submit offers in the day-ahead energy market, clear this market, exactly generate the amount during the scarcity event that they cleared in that market, and which have no surplus capacity beyond what cleared, have no ability to receive the Scarcity Price for energy. They therefore receive none of these incentive revenues. This is an odd construct given that in most markets resources are expected, if not required, to submit offers into the day-ahead market.
3. By design, Scarcity Pricing is tied to one price (or percent of one price) that is assumed to represent the value to consumers for reliable service. The use of one value for the VOLL, regardless of the time of day, time of year, class of customer, and duration of outage is inconsistent with a reasonable understanding of this parameter [23].
4. Up until recently, LOLP has been used as an annual, long-term planning metric. The application in ERCOT and other markets to operating time scales and conditions is misguided. Therefore, unless the markets develop a mechanism to determine an LOLP-type metric given the exact operating conditions in place during the scarcity conditions, there is no foundation for its use in today's markets.
5. Finally, the reliance on the energy markets to provide investment signals is inconsistent with the fundamental marginal cost theory as developed by Dupuit, Hotelling, Coase, et al. Marginal costs cannot be used to incentivize investments when such costs are lower than average costs.

5.2 Critique of Capacity Markets

The Capacity Markets in use in PJM, ISO-NE, NYISO and, to a limited degree in MISO, also suffer from poor design concepts. These include:

- **Short-term market horizon** Thirty- to fifty-year investments cannot be optimized in a market that only provides for one- to seven-year contracts.[11] The consequence is that potential investors demand higher returns on equity due to the uncertain long-term economics of the arrangements (that is, they shift the cost of these risks to the consumers), and they will naturally be biased toward resources that have lower capital costs and higher variable costs (favoring peaking resources over base load resources) [31, Pg. 73].
- **Transmission Investment Coordination** Current market designs provide no explicit co-optimization with potential transmission improvements, nor much, if any, implicit co-

[9] *Operating reserves* represent resource capability above firm system demand required to provide for regulation, load forecasting error, equipment forced and scheduled outages and local area protection.

[10] *Planning reserves* represent installed capacity above the forecasted peak-hour firm system demand for a defined period in the future.

[11] "For the prospective investor in an expensive, forty-year asset, it is next to impossible to estimate the probability that the competitively priced energy produced by the asset will produce a sufficient return over its lifetime (compared to existing or yet-to-be-invented alternatives), or whether the asset will be rendered obsolete or uncompetitive by new regulation [35]."

optimization. Investors look only to existing and potential transmission topologies to decide on resource locations and have little to no control or influence over what the transmission owners may or may not do in the future.

- **Natural Gas Infrastructure Coordination** Like with transmission, the lack of coordination with natural gas investments can significantly hamper the market from realizing the ultimate economically efficient solution.
- **Reliance on Historical Energy Prices to set Capacity Price Caps** – The demand curves developed by each Capacity Market are based on the *Cost of New Entry*[12] for that market, and nets out the expected value of energy revenues that such a hypothetical resource would realize in the market. However, these estimates are based on historical prices (PJM, for example, uses an average of the past three years). Because the demand curve covers a period three years in the future (for PJM), this means there is a six-year difference. Such a difference could mean that future energy market prices could be significantly higher or lower than those assumed for the demand curve.

The historical methods used to plan for and optimize resource expansion are not without faults. Projects were sometimes planned and built to serve load that never materialized. Some projects experienced significant cost overruns, with these costs typically passed on to consumers under the cost-of-service, rate-of-return paradigm.[13] Some of these cost overruns were due to changes in regulations during the development and construction of projects, but some were due to poor management, or worse.

6 Resource Investment Solutions

We believe the solution to the problems identified with today's Capacity Markets lie in one or more hybrid approaches that borrow from the methods perfected over time by traditional resource planners and those methods that utilize competitive markets. These solutions are consistent with the expansion methods applied to industries with similar average-to-marginal cost structures like the airline industry, the automobile industry, and the hotel industry.

Three specific options are summarized below that implement this proposed construct.

- **Laissez Faire** – Allow LSEs to secure sufficient capacity to meet their capacity requirements (defined as serving their load, plus a level of reserves prescribed of the RTO) by any means they deem acceptable, as long as the capacity acquired meets criteria established by the RTO.
- **Long-Term Capacity Markets** – Require each LSE to have a portion of their capacity requirements (perhaps the majority of their requirements) secured for a longer-term period (e.g., 20+ years instead of 1 year).[14] "More forward contracting would be a good thing from both a market power mitigation perspective and from the perspective of those who believe that price volatility, price uncertainty, and opportunism are deterrents to investment [21]."
- **RTO as the System Planner** – Have the RTO plan for the entire market footprint, competitively bid for generation to meet requirements, and allocate costs to market participants.

While we believe the above approaches are all superior to those currently adopted in the capacity markets in the U.S. today,[15] it is the last approach that RTOs should implement. Under this best-of-both-worlds approach, the most valuable features of the traditional expansion planning approach[16] would be retained, while competition for the development, ownership and operation of the resources as seen in today's RTO Capacity Markets would be preserved.

Similar to approaches utilized in CAISO and in the "non-organized" markets in the Western and Southeastern U.S., this approach would involve a process where the following are determined through a market-wide planning process that is conducted by the RTO itself:

- Reliability requirements are established and tracked
- Future load requirements are forecasted
- Planned generation and transmission assets are identified and incorporated
- Long-term analyses are performed by the RTO that identify: i) the amount of capacity needed, ii) the desired location of the capacity considering existing and potential transmission and natural gas (or other fuel supply) infrastructure, and iii) the desired technology of resources used to provide the capacity (supply or demand

[12] The *Cost of New Entry*, or CONE, is an estimate of the cost to build the least-cost resource in each market.

[13] In August 2017, utilities in South Carolina announced the abandonment of the construction of two new units at the Summer Nuclear Station.[32]

[14] Forward, multi-year contracting has long been recognized as a means to mitigate problems experienced in the short-term energy markets [2, 40].

[15] SPP is excluded, because it is essentially the Laissez Faire approach and not a "formal" capacity market per se.

[16] The planning methods and "best practices" used and perfected by utilities over the past century.

side), with proper assessment of risks associated with newer technologies.

- Following development and agreement on a plan for the market, competitive auctions will be held for suppliers to build the desired resources, who would bear construction and performance risks.
- The RTO will contract with the successful bidders for the purchase of capacity and associated energy under long-term (e.g., 20-year+) agreements.
- LSEs will be allocated[17] the cost of capacity required to reliably serve the market based on their load-ratio share.

A competitive process would then follow that would determine who would provide the desired resources and what the final prices for these resources would be.

As suggested by [35], the reason the proposed approaches are superior lie in the foundational questions of energy policy. "How will society manage risk and uncertainty in energy markets?" "How will it manage the distribution of external costs and benefits not captured by market prices?" The proposed approaches will do this.

With the RTO as the overall System Planner, it can incorporate all the "good" from traditional planning experience and take advantage of economies of scale to develop market-wide: (i) load and fuel price forecasts, (ii) technology assessments, (iii) transmission and fuel supply infrastructure studies, and (iv) assessments of political, legal and regulatory frameworks within which that markets may operate in the future. And like the energy markets, these can all be done using best-in-class systems, models and methods.

Critics of the traditional approach to expansion of power systems should also be satisfied. While such plans will be developed by the RTO, construction and performance risks will be borne by the independent developers and owners of the resources. Cost overruns and performance penalties will therefore not be directly passed through to eventual customers.

7 Conclusions

Because, by design, the Capacity Market solutions used by four of the RTO markets in the U.S. fail to incorporate many of the engineering principles used in long-term resource acquisition, they are unable to provide adequate investment signals to existing and potential resource owners [31]. This includes the lack of a contract term consistent with the engineering time-scales associated with generation investments. They also ignore most, if not all, of the strategic aspects concomitant with long-term planning (impacts on transmission, fuel supply infrastructure, fuel diversity, etc.). The solutions provided address these shortcomings.

As described by Hayek [17] "planning" is the "complex of interrelated decisions about the allocation of our available resources." In this context, Hayek believed all economic activity can therefore be viewed as planning. And "in any society in which many people collaborate, this planning, whoever does it, will in some measure have to be based on **knowledge** which, in the first instance, is not given to the planner but to somebody else, which somehow will have to be conveyed to the planner."

The key issue is therefore not that planning is done, but who is to do the planning. So the key dispute regards whether planning is to be done centrally, by one authority for the whole power supply system, or is divided among many individuals.

Challenges in the creation of capacity markets has been the focus of this paper. While short-term energy and ancillary markets are not perfect, the outcome is largely as intended: the RTOs took well-developed engineering methods and approaches used by electric utilities to optimize power supply systems in the short-term, and simply applied them to larger systems. They have also taken advantage of improvements in enhanced computing power to simultaneously optimize energy, reserves and transmission over these larger systems—something not possible until relatively recently.

However, unlike the energy and ancillary services markets, the RTOs ignored most of the long-developed engineering methods and approaches to optimizing power system expansion plans that covered longer time periods. This error adversely impacts the investment incentives for all market participants. In essence, and again unlike the energy markets, the RTOs did not take a well-functioning system and make it better—they made it worse.

References

[1] R. Aïd, S. Federico, H. Pham, and B. Villeneuve. Explicit investment rules with time-to-build and uncertainty. *Journal of Economic Dynamics & Control*, 51:240–256, 2014.

[2] B. Allaz and J. Vila. Cournot competition, forward markets and efficiency. *Journal of Economic Theory*, 59(1):1–16, 1993.

[3] W. J. Baumol and D. F. Bradford. Optimal depar-

[17] One method is to charge each LSE an amount based on their "load ratio share" of the market's total capacity requirement. This is similar to how capacity costs are allocated to LSEs in some RTOs today.

tures from marginal cost pricing. *American Economic Review*, 60(3):265–283, 1970.

[4] R. E. Bohn. A theoretical analysis of customer response to rapidly changing electricity prices. *MIT Energy Laboratory*, 1981.

[5] R. E. Bohn. Spot pricing of public utility services. *MIT Energy Laboratory*, 1982.

[6] R. E. Bohn, M. Caramanis, and F. Schweppe. Optimal spot pricing of electricity. *MIT Energy Laboratory*, 1980.

[7] S. Borenstein. Electricity pricing that reflects its real-time cost. Technical report, NBER Reporter: Research Summary, 2009.

[8] J. Bowring. Capacity markets in PJM. Technical report, Monitoring Analytics, 2013.

[9] J. Bushnell, M. Flagg, and E. Mansur. Capacity markets at a crossroads. Technical report, Energy Institute at Haas, 2017.

[10] Y. Chen, U. Hashmi, J. Mathias, A. Bušić, and S. Meyn. Distributed control design for balancing the grid using flexible loads. In *IMA volume on the control of energy markets and grids*. Springer, 2017.

[11] I. K. Cho and S. P. Meyn. Efficiency and marginal cost pricing in dynamic competitive markets with friction. *Theoretical Economics*, 5(2):215–239, 2010.

[12] A. J. E. J. Dupuit. De la mesure de l'utilite des travaux publics. *Annales des ponts et chaussees*, 2(8), 1844.

[13] M. Feldstein. Equity and efficiency in public sector pricing: The optimal two-part tariff. *Quarterly Journal of Economics*, 86, 1972.

[14] FERC. *Guide to Market Oversight: Glossary.*

[15] Florida Power & Light. Ten Year Power Plant Site Plan. Technical report, 2016.

[16] R. R. Geddes. *Historical Perspective on Electric Utility Regulation.* CATO Review of Business and Government, 1992.

[17] F. A. Hayek. The use of knowledge in society. *The American Economic Review*, XXXV(4):519–530, 9 1945.

[18] W. W. Hogan. Electricity scarcity pricing through operating reserves. *Economics of Energy & Environmental Policy*, 2(2):65–86, 2013.

[19] H. Hotelling. The general welfare in relation to problems of taxation and of railway and utility rates. *Econometric Society*, 1937.

[20] H. Hotelling. The relation of prices to marginal costs in an optimum system. *Econometrica*, 7(2):151–155, 1939.

[21] P. L. Joskow. Competitive electricity markets and investment in new generating capacity. pages 1–74, 2006.

[22] P. L. Joskow. Symposium on 'Capacity Markets'. *Economics of Energy and Environmental Policy*, 2(2):v–vi, 2013.

[23] D. Keane. Outage cost estimation guidebook: EPRI Research Project 2878-04 Final Report. Technical report, EPRI, December 1995.

[24] A. P. Lerner. *The Economics of Control: Principles of Welfare Economics.* Macmillan, 1944.

[25] A. Marshall. *Principals of Economics.* Macmillan and Co., Ltd., 1890.

[26] A. Mas-Colell. *Microeconomic Theory.* Oxford University Press, 1995.

[27] J. E. Meade. Price and output policy of state enterprise. *Economic Journal*, 1944.

[28] M. Negrete-Pincetic. *Intelligence by design in an entropic power grid.* PhD thesis, University of Illinois at Urbana Champaign, University of Illinois, Urbana, IL, USA, Dec. 2012.

[29] M. Negrete-Pincetic and S. Meyn. Markets for differentiated electric power products in a Smart Grid environment. In *IEEE Power and Energy Society General Meeting*, pages 1–7, 2012.

[30] J. R. Nelson, editor. *Marginal Cost Pricing in Practice.* Prentice-Hall International, 1964.

[31] M. Petitet. *Long-term dynamics of investment decisions in electricity markets with variable renewables development and adequacy objectives.* PhD thesis, 2016.

[32] B. Plumer. U.S. Nuclear Comeback Stalls as Two Reactors Are Abandoned. *New York Times (Online)*, July 2017.

[33] T. Schroder and W. Kuckshinrichs. Value of Lost Load: An efficient Economic Indicator for Power Supply Security? A Literature Review. *Frontiers in Energy Research*, 3, December 2015.

[34] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn. *Spot Pricing of Electricity.* Kluwer Academic Publishers, 1988.

[35] D. B. Spence. *Naïve Energy Markets.* Springer, 2017.

[36] R. D. Tabors and J. L. Kirkley. Homeostatic control: The utility/customer marketplace for electric power. *MIT Energy Laboratory*, 1981.

[37] R. Treinen. Locational Marginal Pricing (LMP): Basics of Nodel Price Calculations. CAISO Market Operations Presentation, December 2004.

[38] R. Turvey. Marginal cost pricing in practice. *Econometrica*, 31(124):426–432, November 1964.

[39] G. Wang, M. Negrete-Pincetic, A. Kowli, E. Shafieepoorfard, S. Meyn, and U. V. Shanbhag. Dynamic competitive equilibria in electricity markets. In A. Chakrabortty and M. Ilic, editors, *Control and Optimization Methods for Electric Smart Grids*, pages 35–62. Springer, 2012.

[40] F. Wolak. Quantifying the supply-side benefits from forward contracting in wholesale electricity markets. *Journal of Applied Econometrics*, (22):1179–1209, 2017.

[41] F. Wolak. Capturing the benefits of California's energy infrastructure investments. Technical report, Stanford University, Accessed January 2012.