The Efficiency of Multi-Unit Electricity Auctions*

Wedad Elmaghraby† and Shmuel Oren‡

October 4, 1998

Abstract

Using a complete information, game-theoretic model, we analyze the performance of different electricty auction structures in attaining efficiency (i.e., least-cost dispatch). We find that an auction structure where generators are allowed to bid for load "slices" outperforms an auction structure where generators submit bids for different hours in the day.

1 Introduction

The electric power industry around the world is undergoing a process of privatization deregulation and restructuring. This transition is fueled by technological and social change that led to a fundamental reexamination of conventional wisdom concerning natural monopolies and economies of scale in this industry.

While the restructuring approaches implemented or proposed in various parts of the world and within the US are diverse in many aspects they share several important elements which include competitive generation, a spot energy market and a power auction. The purpose of the auction is to provide a mechanism through which generators can submit bids to supply electricity. The most challenging aspect of designing an electricity auction is that daily demand, which fluctuates from period to period, must be satisfied by a set of suppliers, with different cost, in a least-cost manner. Even in a centralized model with known generator costs determining the optimal dispatch is a computationally difficult problem. It is an even greater challenge to design an auction where generators voluntarily chose to be efficiently dispatched.

In any electricity auction, generators must submit bids which indicate the minimum prices at which they are willing to generate electricity. In some auction designs (e.g. the in the UK system) bids can also include state transition costs such as start up and ramping costs as well as constraints on availability and dispatch[1]. The structure of the auction and the determination of prices paid to the winning bidders can vary. It is desirable that the designer of an electricity auction define the structure and prices in such a way as to provide generators with the incentive to bid so as to minimize generation costs. We shall refer to the set of generators which minimizes generation costs as the *efficient dispatch* and to an auction which induces an efficient dispatch as an *efficient auction*.

There are several aspects of an auction for electricity that separate it from the vast body of auction literature and make designing an efficient auction a challenging task. The most obvious is the structure of generation costs. Generators have many cost components (e.g., ramp-up costs, no-load costs, etc.) which

*An earlier abridged version of this paper appeared in the conference proceedings of the IAEE Fall 1997 meeting in San Francisco.

†Stern School of Business, New York University

‡Dept. of IEOR, University of California at Berkeley

[1] See Patrick and Wolak (1996) for an analysis of the United Kingdom auction design.

must be recovered through their sales revenue. In addition generators are subject to intertemporal dispatch constraints that relate their output in different time periods. These characteristics create cost dependencies in *intertemporal* production so that the average cost of generating Q GW of electricity varies with the number of units generated and the dispatch schedule. Such dependencies complicate both the bidding strategies and the bid evaluation protocols.

The long standing tradition of vertical integration and centralized dispatch in the electric utility industry resulted in advanced computational tools for optimal dispatch of generating resources, which take as inputs all the costs and operational constraints for each available resource as well as demand data and reliability requirements. Such tools are designed based on the premise of perfect information about these inputs and produce as outputs two type of decision variables. The optimal commitment schedules specifying the periodly on/off state of each resource are typically produced for a range of the next 168 periods. These schedules and a rough estimate of the periodly output level of each resource are calculated by a unit commitment algorithm which is a mixed nonlinear and integer optimization program run every period on a rolling horizon basis for the next 168 periods. The unit commitment schedules are used as inputs to an optimal power flow calculation which uses a nonlinear optimization algorithm run repeatedly at short time intervals to obtain the up-to-the-minute output levels of each generator on line. The optimal power flow employs a more realistic model of the power system that takes into consideration transmission and security constraints as well as various physical aspects associated with alternating current (AC) systems.

Some restructuring designs have attempted to preserve the central unit commitment protocols with competitive generation by employing a multipart auction to elicit the inputs needed for the traditional unit commitment algorithms. In such auctions bidders are required to submit for the day ahead, supply functions for energy as well as all the other cost components and dispatch constraints needed for central unit commitment procedures. Such an auction structure is employed in the UK where unit commitment is performed using the GOAL algorithm and it is part of the proposed restructuring plan for the New York Power Pool. Unfortunately multi-part auctions are not well understood and have limited theoretical foundation that would enable an incentive compatible design of such auctions. Indeed, the UK experience suggests that such auctions are susceptible to gaming and manipulation. Furthermore, recent work by Johnson, Oren and Svoboda (1997) suggest that even if such an auction was made incentive compatible (i.e., bidders would be induced to reveal true costs and constraints) central unit commitment may still be inappropriate in a competitive generation environment. In particular, the authors have demonstrated that unit commitment algorithms designed for an environment with central generation ownership have multiple equally good solutions with varying resource schedules. When generation ownership is dispersed among many independent parties, such variation have diverse profit implications for the different parties. The resulting ambiguity in the bid selection protocols cannot be resolved by tie breaking procedures since it is not practical to compute all the optimal solutions (even one good solution is computationally challenging). Furthermore, the optimal unit commitment produced by a specific algorithm (out of the many possible) is affected by fine tuning of the program which consequently may be systematically biased in favor of some generators to the detriment of others.

An alternative approach to the multi-part auction that has been adopted in the California Power Exchange protocol and in the Victoria pool in New Wales, Australia relies on self-commitment. In other words, unit commitment decisions are left to the bidders while the auction structure is simplified to a single price per tender per plant.[2] A tender consists of one or multiple blocks of energy defined in terms of their timing and capacity (e.g. 2 GW supplied for one period between 1 and 2 PM). All the production costs incurred by a generator (including fixed, intertemporal and dispatch constraints costs) are internalized in such an auction and reflected in the single bid price.

From an auction theory perspective power auctions with self-commitment may be interpreted as multi-unit auctions with complementarities. There are a few papers that address the issue of multi-unit auctions.

[2] In reality, generators may own several units (gensets) and may submit a separate bid per unit. I shall refer to a unit as a plant.

Wilson (1979) began the study of "share" auctions, where bidders with a common valuation submit demand curves and are awarded a fraction of a unit's shares at a market clearing price. He concludes that a seller's revenue is greatly reduced using a share auction when compared to selling the entire unit as an indivisible object. Maskin and Riley (1989) study the design of optimal multi-unit auctions with private valuations and show that neither a uniform price or discriminatory price auction is an optimal auction mechanism, i.e., neither maximize the seller's expected revenue. Hausch (1986) studies a two object auction, where there are two bidders with common valuations who desire both objects. Hausch finds that the seller's revenue is greater when both objects are sold simultaneously versus sequentially. Rothkopf, Pekec, and Harstad (1995) identify a special class of multi-unit auctions in which bidders can submit a bid for different combinations of objects and the auction is computationally manageable.

Bikhchandani (1996) establishes the allocative efficiency properties of first[3] price auctions when several heterogeneous objects are sold simultaneously (one auction for each type of object) and bidders may desire more than one object. The bidders' are assumed to have no budget constraints, private valuations for the objects that are common knowledge. Bikhchandani finds that, for a first price auction, an efficient pure-strategy Nash Equilibrium exists if and only if a Walrasian equilibrium exists. This result holds whether there is one or several of each type of object sold. In the case of several of each type of object, the identical objects are sold in the same auction and the bidders are allowed to submit a separate bid for each of the identical objects.

There has been as recent growth of auction theory literature in this area in response to the recent FCC spectrum auctions and the multitude of interesting questions they have raised. In the FCC auctions, bidders, comprised of US telecommunication companies, cellular telephone companies, and cable-television companies, competed to win various spectrum licenses for different geographical area. The synergies arising from owning licenses in adjoining geographical area create dependencies in (some) bidders' valuations for individual licenses (see McMillan (1994), Cramton (1995) and McMillan and McAfee (1996) for further discussion of the FCC spectrum auctions). Ausubel and Cramton (1996) question the superior allocative efficiency properties of uniform pricing rules using Wilson's (1979) "share" auction framework with private valuations. They find that the efficiency of 2^{nd} price (uniform) auctions in a single-unit auction do not carry over to a multi-unit framework. They conclude that when bidders desire more than one object, or a large share of the total objects being auctioned, they have an incentive to underbid or "shade" their bids, resulting in an inefficient allocation. Milgrom (1998) explains the different auction formats that were candidates for the FCC spectrum auction and their relative strengths and weaknesses. Isaac and James (1998) use an experiment setting to evaluate the performance of a Vickrey auction in a simple setting with multiple units which exhibit synergies in value.

Von der Fehr and Harbord's (1993) analysis of the UK electricity industry is the only other study we know of that identifies an electricity auction as a multi-unit auction with private valuations and attempts to study the strategic bidding behavior of generators. Von der Fehr and Harbord assume a two generator framework with uncertain demand and known marginal costs. They show that the less efficient (higher marginal cost) generator may submit lower bids than the more efficient generator, and hence generation costs may not be minimized in equilibrium.

Our objective in this paper is to address specific types of such auctions that are relevant to the context of electric power. In particular we focus on three alternative ways of structuring a multi-unit power auction and use the framework of games with perfect information to examine the efficiency of their outcome. Specifically, we focus on the following question: Will the proposed auction structure induce the efficient (i.e., least social cost) dispatch in a non-cooperative[4] setting where generators know each others' costs? While the analysis employs a very simplistic stylized model of demand and generation cost, we employ the insight of the analysis

[3] A first price auction represents a first-price sealed bid auction.

[4] Given the fact that deregulation is often being coupled with divestiture of generation plants (as evidenced in California) it is plausible to assume that with several companies participating in the market place, collusion will be difficult to sustain despite the repetitive (daily) nature of the auction.

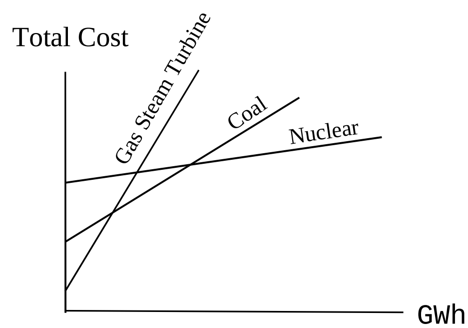

Figure 1: Costs for different generation technologies.

to outline the practical implementation of a new auction structure that promises efficient self-commitment and dispatch.

2 Electricity Market

2.1 Bidders

In an electricity auction, generators compete in price of generation to win dispatch. Generation technologies, for example, nuclear, combined-cycle gas turbines plants, and combustion gas turbines, are generally characterized by four traits. First, a generator's costs to generate fall into two general categories; there is a fixed "start-up" cost incurred each time a generating plant is turned on, and variable cost per GWh once the plant is up and running. Second, there exists an inverse relationship between the start-up cost associated with a technology and its variable cost. For example, a nuclear plant has a large start-up cost but relatively small variable cost per GWh, while a gas-steam turbine has a relatively low start-up cost, but incurs a large variable cost per GWh. A third trait is that generation plants have a constraint on the maximum number of GW they generate at any point in time and are unable to store electricity,[5] but have few restrictions on the duration for which they can generate.[6] Finally, the size of a generation plant is generally decreasing in its variable cost, i.e., the capacity of a low variable cost nuclear plant is much larger than that of a high variable cost combustion gas turbine plant.

For expositional purposes only,[7] we will focus our attention on generation technologies that are efficient, i.e., least-cost, for some output level. Low start-up, high variable cost technologies are the most efficient source over low output (total GWh) levels, while high start-up, low variable cost technologies are the most efficient source over higher output levels. Figure 1 plots the total cost of generation associated with different technology types, assuming a generating plant is "switched-on" only once per day. The horizontal axis measures the total number of GWhs generated over time.

[5] Some generating plants are able to store the *potential* for generating electricity, e.g., hydroelectric generators can store water, but generators are unable to store electricity. Therefore, there will always be a limit on the total MW a generating plant can generate.

[6] Generators do occasionally have to go off-line for maintenance, but this is not a relevant constraint over one day.

[7] The existence of inefficient technologies will not change the results of this paper.

2.2 Market Structure

An auction for electricity must assure that the demand over an entire day will be satisfied by a set of generators. Electricity demand varies through out the day, with highest demand during the afternoon periods and the lowest demand during the middle of the night. Figure 2 plots the total daily demand in California during the middle of March, June, September and December. Generally, demand is described to be one of three types : base, shoulder or peak load.

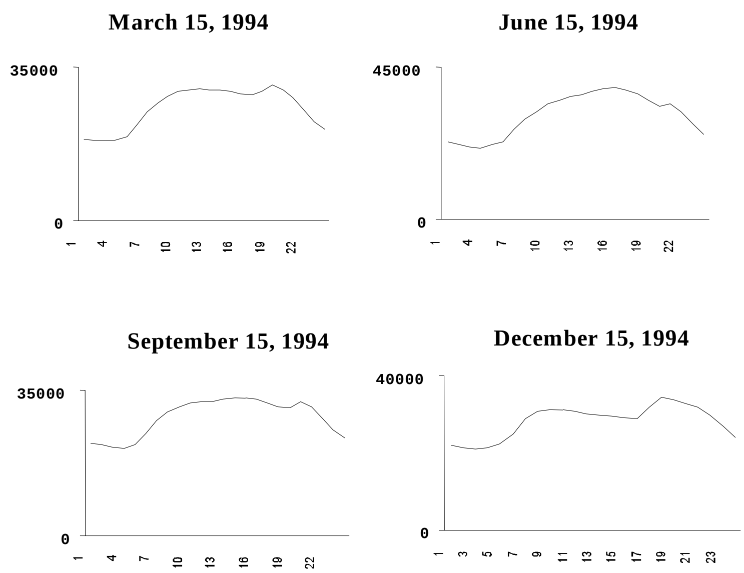

Figure 2: California demand load during the months of March, June, September and December in 1994.

Several types of generation sources are needed in order to meet the varying consumption of demand throughout the day. Generation technologies can be classified by three types; base (b), shoulder (s) and peaking (p). Examples of base, shoulder and peaking plants are nuclear, coal and combustion gas turbine, respectively. Typically a generating plant's capacity, i.e., the maximum rate at which it can generate at any point in time, is a decreasing function of its variable cost and an increasing function of its start-up cost. Nuclear plants, which have high start-up costs, are classified as base plants and typically have a capacity several times larger than a shoulder plant such as a coal plant, which in turn has a larger capacity than a peaking plant such as a gas-steam plant. In the California market, the average (height of) base load, shoulder, and peak demand is approximately 21,10, and 5 GW respectively, while the average size of a base load, shoulder, peak plant is 2,1, and 0.5 GW respectively. This fact necessitates several generation plants of each type to be turned on in order to satisfy demand in a least-cost manner.

2.3 Auction Structure

In this paper we will ignore transmission constraints and assume that the power auction treats all the demand and supply as if it was at a single location. This simplification is consistent with the UK system, the California power exchange, the Victoria pool and other systems around the world where transmission constraints and congestion management are handled outside the power auction. We will further assume that there is no demand side bidding which is also consistent with most currently operating and proposed power auctions. Under this simplified structure the objective of the power auctioneer is to "fill" a forecasted load curve for a specified time period (say the next 24 periods) with tenders consisting of blocks of energy specified by capacity (GW) duration (periods) and timing (see Figure 3(a)).

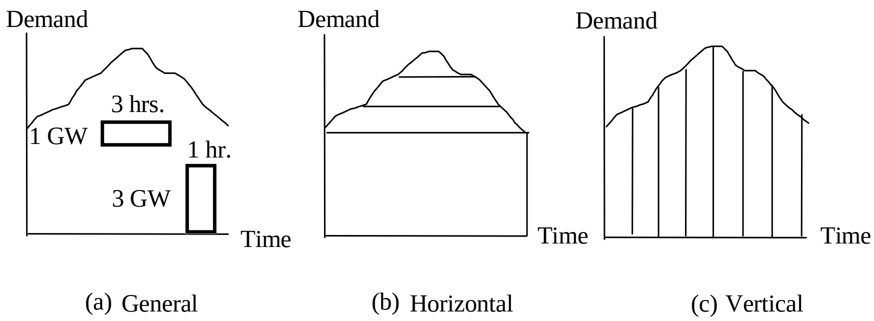

Figure 3: Various partitioning forms of demand load for auctions.

There are many possible ways to structure such an auction in terms of the tenders allowed, the bid evaluation process and the prices paid to winning tenders. In this paper, we will analyze the performance of three types of auction structures, a Daily Supply Curve-vertical (DSC-vertical), a Hourly Supply Curve-vertical (HSC-vertical)[8] and a horizontal auction, and examine their ability to yield efficient dispatches. These auctions can be viewed as combinatorial auctions where the combinations are restricted to be one of two forms, either vertical or horizontal.

In all of these auctions, generators submit (convex) step supply curves. The supply curves indicate the minimum price that must be paid to generate, with each step corresponding to one generation plant (see Figure 4). From the submitted bids, the auctioneer constructs a cumulative supply curve which is used to dispatch generators.

In a *vertical* auction (see Figure 3(c)), daily demand is partitioned into hourly (or half-hourly in the case of the UK) markets. Generators submit supply curves that state the minimum price at which they are willing to generate. All generators chosen for dispatch are paid the highest accepted bid price. The decision to schedule a generator in any period t is determined solely on its bids for that period and independently of its schedule in any other period.

The electricity auctions in the United Kingdom and California can both be characterized as *vertical* auctions. A DSC-vertical and HSC-vertical auction differ in the number of supply curves the participating generators are allowed to submit. Using the characterization set forth in Von der Fehr and Harbord (1993) and Green and Newbery (1992), in the UK electricity auction, generators are allowed to submit a *single* supply curve, which is valid for the 24-hours covered by the auction, indicating the minimum price at which they are willing to generate at different output levels.[9] The auctioneer then constructs a single cumulative supply curve that is used to dispatch the generators for all periods in the day. In contrast, in the California

[8] These auction structures are abstractions from the auctions currently in practice. A daily supply curve (DSC) is used in the England and Wales electricity auction. Likewise, the California Power Exchange accepts bids for each hour in the day. However, the auctions addressed in this paper are not exact characterization of the actual operating auction. For example, a DSC auction differs from the England and Wales auction in that a DSC auction has a uni-dimensional, price only bid while the bid structure in England and Wales is multi-dimensional (with a capacity availability component amongst other operating characterisitics).

[9] In Von der Fehr and Harbord (1993) the supply curves are assumed to be step functions while Green and Newbery(1992) assume that generators submit smooth supply curves.

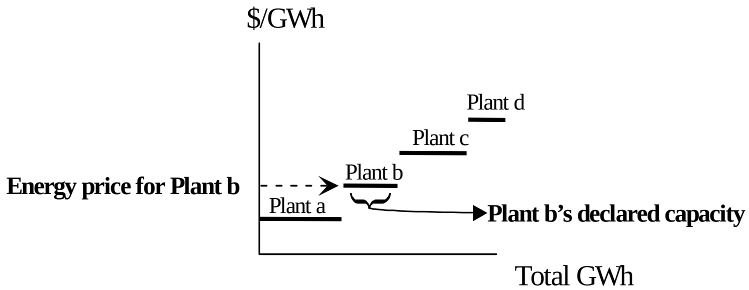

Figure 4: Structure of energy bids; separate bid per generating plant.

electricity auction, generators are allowed to *simultaneously* submit a separate supply curve for each hour in the day. All generators chosen for dispatch in an period are paid the marginal price in that period, i.e., the highest accepted bid price. We will show that there are inherent problems with trying to bid non-convex generation costs into a vertical auctions, regardless of the number of bids allowed.

An alternative auction structure which, we will argue, lends itself more readily to the intertemporal cost dependencies in generation is a *horizontal* auction. In a *horizontal* auction (see Figure 3(b)), demand is divided into distinct horizontal demand sets that are auctioned *sequentially.* Demand sets are formed by grouping demand roughly according to its duration, i.e., a distinct set for each duration t. Hence, generators submit a supply curve for each set, indicating the price at which they are willing to generate k megawatts for a **duration** of t periods, where $k \in (0, K)$ for a fixed $t > 0$. The auctioneer constructs a least-cost supply curve from the submitted bids in each set and the winning bidders are *paid their bid price.* The sequencing of the auctions is such that the demand set with the longest duration is auctioned first, then the second longest, and so forth. The results of any previous auctions are made known before each auction.

One interpretation of a horizontal auction is that the auctioneer is segmenting demand by its types (base, shoulder or peak), thereby creating distinct auctions for the different types of demand. Generating electricity for the 24-period base load demand is a fundamentally different product than generating for just the few peak periods. By partitioning demand by its types, generators are able to submit separate bids for the different types of products provided.

The intuitive motivation for the alternative bid formats is articulated by Wilson (1998) who notes,

> "The bid format is a key factor. For example, if the market is organized to provide hourly schedules and prices, then this tends to serve the interests of demanders for whom the time of power delivery is important, and suppliers with flexibility (e.g., hydro), whereas it tends to ignore the considerations of suppliers from thermal sources, who are mainly concerned with obtaining operating schedules over consecutive hours sufficient to recover the fixed costs of start-up and who are unconcerned about timing *per se*. Schemes have been devised that allow demanders to bid on a time-of-day basis while suppliers bid for operating runs of various durations; prices can then be stated equivalently in terms of hourly prices for demanders and duration prices for suppliers."

In the next section, we assume a simple model for demand and generation costs and demonstrate why neither a DSC-vertical or HSC-vertical auction can guarantee the efficient dispatch in equilibrium. We then prove that a horizontal auction *does* guarantee efficiency in our model. We conclude with the reasons for and the intuition behind these results.

3 Model

For simplicity, we assume that demand over the length of a day is represented by a step function (Figure 5(a)). Figure 5(b) represents the daily demand in terms of its load duration curve. When daily demand is single peaked, we can use the load duration curve in our analysis without loss of generality. Electricity auctions are conducted on a day ahead basis, where generators bid to supply the next day's forecasted demand. We model the daily demand as public information and known to the generators. In particular, we will assume that the load duration curve is as in Figure 6, i.e., there is demand during only three periods of that day and demand is constant within each period.

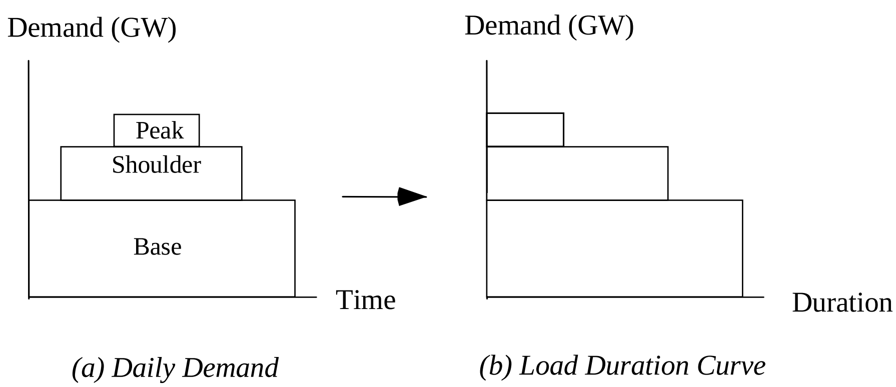

Figure 5: Daily demand represented as a load duration curve.

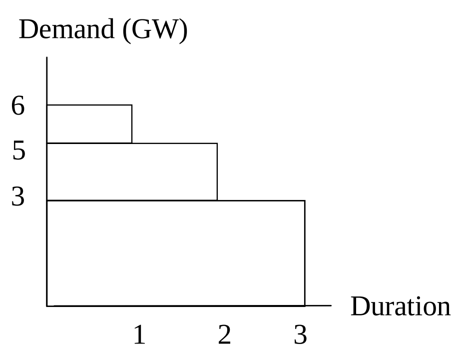

Figure 6: Assumed demand load.

Throughout the paper we assume that generators have two costs, which are publicly known, associated with generation; a fixed "start-up" cost, f, to begin to generate and a variable cost per GWh, v, once the plant is up and running. Due to this cost structure, there exist cost dependencies in *intertemporal* production. In the following sections, we assume that there are three types of generation technologies; base

Technology Type	Capacity
Base	3
Shoulder	2
Peak	1

Table 1: Capacities of technology types

(b), shoulder (s) and peaking (p) (see Figure 7), where a generator of technology x is denoted by G_x. We assume that there exist many generation companies participating in the market, where all the plants owned by a generation company are of the same technology type. In the main body of this paper, for expositional purposes only, we further assume that each generation company owns only one plant. However, our analysis and results carry over to a framework where generation companies own many plants of the same technology type and, in the extreme case, when each generation company is the sole owner of a particular generation technology.

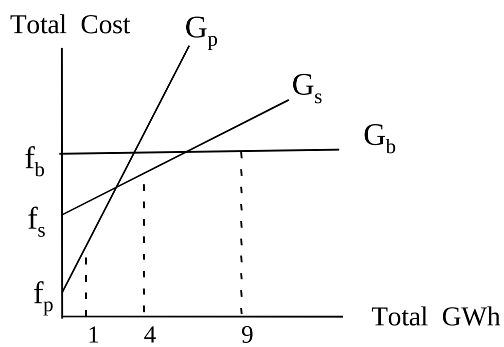

Figure 7: Assumed generation types and respective costs.

Table 1 lists the assumed capacity constraints for each type of plant. The capacity limitations imply that the maximum a base (shoulder, peak) load plant can generate at any point in time is 3 (2,1), but imposes no constraint on the duration for which a plant can generate.

Under these assumptions, we are exploring the simplest scenario where the size of the base(shoulder, peak) demand is equal to the capacity of one base(shoulder, peak) plant. The capacity plus demand assumptions imply that there is a unique efficient dispatch, as given in Figure 8. This simple example is rich enough to capture the failings and strengths of the three auctions structures presented here. In the next section we shall see that neither form of a vertical auction is able to guarantee the efficient dispatch in equilibrium.

3.1 Vertical Auctions

3.1.1 DSC-vertical Auction

In a vertical auction where generators own only one plant, each generator submits a supply curve which consists of one point: the minimum energy price at which the generator is willing to generate up to its plants' capacity. The auctioneer then dispatches generators in increasing order of bids. We have found that even in our simple model of cost and demand, a DSC-vertical auction does not support an efficient dispatch in its set of Nash Equilibria.

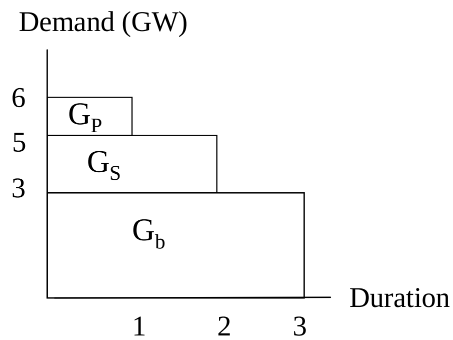

Figure 8: Efficient dispatch for assumed demand load.

Generator	Period 1 π	Period 2 π	Period 3 π	Total Cost
G_b	$3a$	$3(a-\varepsilon)$	$3(a-2\varepsilon)$	f_b+9v_b
G_s	$2a$	$2(a-\varepsilon)$	0	f_s+4v_s
G_p	a	0	0	f_p+v_p

Table 2: Revenues and costs for generators in efficient dispatch under a UK-vertical auciton

Proposition 1 *In a complete information setting, a DSC-vertical auction cannot guarantee the efficient dispatch in equilibrium.*

Proof. Suppose, without loss of generality, that there exists one generator of each type, labeled G_p, G_s, and G_b who submit bids of a, b, and c, respectively. Given a daily demand as in Figure 8, an efficient dispatch requires $c < b < a$ to hold, which implies G_b is dispatched all three periods at his capacity of 3 GW, Gs is dispatched in the first two periods at her capacity of 2 GW and G_p is dispatched in the first period at his capacity of 1 GW. We argue that such bids can not occur in equilibrium as a result of two conflicting issues: The bidders' desire to increase their bids (and hence their revenue) as much as possible while still ensuring dispatch and the discontinuity of their profits as a function of bids in combination with their non-convex generation costs.

Suppose we have $c < b < a$ as required in an efficient dispatch, which results in the cumulative supply curve given in Figure 9. The auctioneer uses the (same) cumulative supply curve to dispatch the generators in all three periods. The intersection of demand in any period with the cumulative supply curve determines the clearing price for that period.

If $c < b < a$ constitute an equilibrium set of bids, it must be that $c + 2\varepsilon = b + \varepsilon = a$.[10] If not, then at least one of the generators has an incentive to increase its bid, and hence its profit in the period in which it is the price-setter, without altering its dispatch schedule. The clearing prices in periods $1, 2$, and 3 would be $a, a-\varepsilon$, and $a - 2\varepsilon$ respectively. In addition, a must be greater than or equal to $f_p + v_p$ in equilibrium. Each generators' revenue in each period and total cost in the efficient dispatch are summarized in the Table 2 below.

The bid ordering $c < b < a$ is an equilibrium set of bids if and only if no generator has an incentive to deviate from its bid. However, G_p does have the incentive to change his bid given $c + 2\varepsilon = b + \varepsilon = a$. Given

[10] We assume that the smallest bid increment is ε.

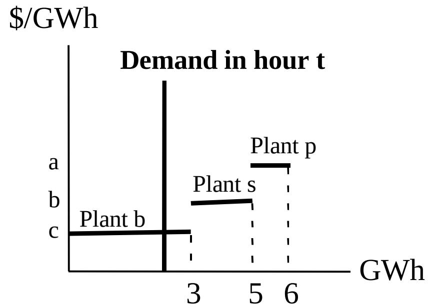

Figure 9: Necessary bids for efficient dispatch to result in a *DSC-vertical* auction.

these bids, G_p would prefer to bid $a-3\varepsilon$ and undercut G_b as the base load generator as long as f_p is greater than two times the smallest bid increment (in this case, as long as $f_p > 2\varepsilon$). By doing so, he foregoes ε in revenue in the first period, but wins a dispatch in the second period at G_s's bid price of $a-\varepsilon$ and in the third period at G_b's bid price of $a-2\varepsilon$. To see why this is true, note that G_p has the incentive to deviate from a bid of a as long as

$$\begin{aligned} a-(f_p+v_p) &< [(a-\varepsilon)+(a-\varepsilon)+(a-2\varepsilon)]-(f_p+3v_p) \\ a &> v_p+2\varepsilon \end{aligned}$$

At its lowest possible equilibrium value, $a = f_p + v_p$, therefore, all that is needed in order for G_p to have the incentive to undercut G_b's bid is for $f_p > 2\varepsilon$. Start-up costs typically run several orders of magnitudes larger than the minimum bid increment (which is 1 cent in the California Power Exchange). Therefore the bid ordering $c < b < a$ will not be supportable in equilibrium and we have shown that the efficient dispatch cannot be supported and hence cannot be guaranteed in equilibrium. ■

This result easily extends to a framework with many generators of each type and therefore is not dependent on the presence of market power.

3.1.2 HSC-vertical Auction

The inability of a DSC-vertical auction to guarantee the efficient dispatch might be thought to be a result of the restriction of one bid per day. The auction in California, in contrast to that in the UK, allows generators to submit a separate supply function for each hour in the day. We shall show that despite the added flexibility of separate bids for each period, the HSC-vertical auction cannot guarantee efficiency in equilibrium. This is not to say, however, that it cannot support the efficient dispatch; both inefficient and the efficient dispatch are supportable in equilibrium.

Proposition 2 *In a complete information setting, a HSC-vertical auction cannot guarantee the efficient dispatch in equilibrium.*

Generators	Bid in Period 1	Bid in Period 2	Bid in Period 3
G_b^1	0	$\frac{f_b+5v_b-2(f_p+v_p-\varepsilon)}{3}$	$f_p + v_p - \varepsilon$
$G_b^i,\ i = 2,3$	$f_p + v_p + \varepsilon$	$\frac{f_b+5v_b-2(f_p+v_p-\varepsilon)}{3} + \varepsilon$	$f_p + v_p$
G_s^1	0	0	0
$G_s^i,\ i = 2,3$	$f_s + 2v_s$	$f_s + 2v_s$	$f_s + v_s$
$G_p^i,\ i = 1,2,3$	$f_p + v_p$	$f_p + v_p$	$f_p + v_p$

Table 3: Inefficient Equilibrium Bids

Proof. Suppose, without loss of generality, that there exists three generators of each type, labeled G_p^i, G_s^i, and G_b^i, $i = 1,2,3$ who submit bids for periods $t = 1,2,3$ respectively. Given the demand in Figure 6 and capacity assumptions, Table 3 defines a set of equilibrium bids for the generators which constitutes an *inefficient* dispatch in the HSC-vertical auction (see Figure 10 for the inefficient dispatch).

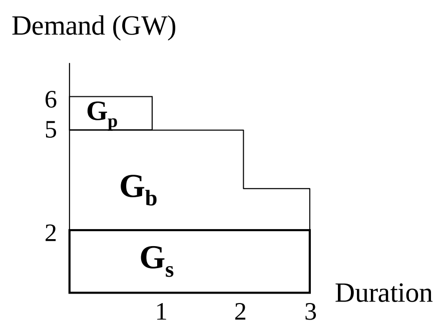

Figure 10: Inefficient dispatch suppoartable in a *HSC-vertical* auction.

Given their opponents' strategies in Table 3, no generator has an incentive to deviate from its bids. This profile of bids results in G_s^1 being dispatched at her capacity of 2 GWh in all three periods, G_b^1 being dispatched at his capacity of 3 GWh in periods 1 and 2 and at 1 GWh in period 3, and G_p^1 being dispatched for 1 GWh in period 1. The clearing price received by all winning generators in period 1 is $f_p + v_p$ per GWh, $f_b + 3v_b - 2(f_p + v_p)$ per GWh in period 2 and $\frac{f_b}{2} + 2v_b$ per GWh in period 3. Since G_s^1's payoff is not determined by her bid, she has every incentive to bid as low as possible to ensure dispatch. By submitting a bid of zero in all three periods, G_{21} is able to win dispatch at a positive profit. Although G_b^1 is (one of) the least-cost producers of 9 GWh, he is unable to profitably undercut G_{21}'s bids of zero. ∎

This simple example clearly illustrates why a HSC-vertical auction cannot guarantee efficiency in an environment with non-convex costs. It is quite likely that in any given period, not all the generators will be dispatched at the same output level. In such a scenario, there exists an opportunity for a relatively inefficient generator to accrue a positive profit by bidding zero and ensuring dispatch without fear of receiving its below-cost bid price. With the knowledge that in equilibrium the clearing price is guaranteed to be at least the cost of the marginal bidder in period t, a relatively inefficient generator can "sneak-in" to the dispatch schedule by submitting a zero bid, get dispatched at a higher level in period t than the marginal price-setting bidder and accrue a positive profit due to the non-convexity of its cost curve. The presence of excess generation capacity for each type reassures us that the *HSC-vertical* auction's failure to guarantee efficiency is not due to market power, but instead points to a more fundamental auction design flaw.

3.2 Horizontal Auction (HA)

We can learn from the failure of a vertical auction to guarantee efficiency in equilibrium that it is necessary to account for cost dependencies in electricity generation when designing an auction. An auction structure which does exactly that is a horizontal auction. Below we provide a simple example which illustrates the intuition behind the ability of a horizontal auction to guarantee efficiency. While the example presented here is designed such that there is only one winner per auction, the results carry over to a more general framework where there is more than one winner per auction.

Assume the same framework of demand, costs and generators (Figures 7 and 8) as in previous section. In a horizontal auction, base load demand is auctioned first. After the winner is made public knowledge, the shoulder load is auctioned and then the peak. In order to attain the efficient dispatch, a G_b technology must win in the base load auction, a G_s technology must win in the shoulder load auction and a G_p technology must win in the peak load auction. Due to the sequential nature of the auctions, the appropriate equilibrium concept to focus our attention on is Subgame Perfect Nash Equilibrium (SPNE).

Proposition 3 *The unique SPNE outcome of the horizontal auction of demand load in Figure 8 is efficient.*

Proof. To prove our claim, we will employ backward induction to identify the Subgame Perfect Nash Equilibrium (SPNE) for the entire game. At the start of auction for the peak load, we can be in any one of nine possible states (see Figure 11 for an exhaustive description of the game and Table 4 for a complete list of the nine possible states at the start of the peak load auction.)

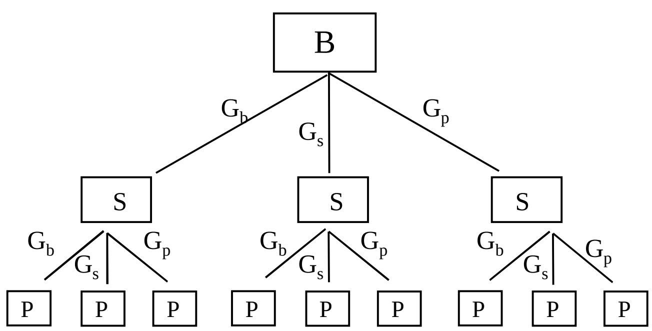

Figure 11: Game tree in a sequential horizontal auction (B =baseload, S=shoulder, P=peak).

Given that there exist a surplus of generators for each technology type, the unique equilibrium is for a G_p generator to win in the peak load auction regardless of what state we are in at the start of the peak load auction. This is because, there will be at least one of each type of generator types participating in the peak load auction. A type G_p generator is the least-cost supplier of the peak load, and therefore will win. Likewise, regardless of which type of generator has won in the base load auction, the unique equilibrium outcome in the shoulder load auction is for a G_s generator to win. As a result, the unique outcome in the base load auction is for a G_b generator to win (unique equilibrium outcome is given in Figure 12).

■

The above analysis can be extended to a framework with n generators of each type where there are more than one winner per auction. In order for the results carry to a more general framework, the size of the

State	Winner in base load auction	Winner in shoulder load auction
1	G_b	G_b
2	G_b	G_s
3	G_b	G_p
4	G_s	G_b
5	G_s	G_s
6	G_s	G_p
7	G_p	G_b
8	G_p	G_s
9	G_p	G_p

Table 4: Possible states at the start of the last auction

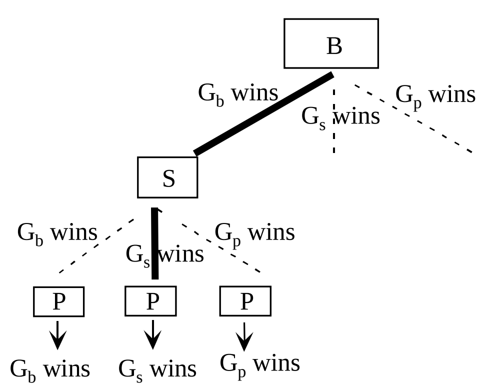

Figure 12: The unique SPNE for the *horizontal* auction.

generators must be small relative to the demand for which they are bidding - which can be reinterpreted as the capacity of a generator is exhausted within an auction if chosen for dispatch. This assumption is well justified in the California market, where the average height of base load, shoulder, and peak demand is approximately 21,10, and 5 GW respectively, while the average capacity of a base load, shoulder, and peak plant is 2,1, and 0.5 GW respectively. Under these assumptions we have proved that all Subgame Perfect Nash Equilibria of a horizontal auction results in the unique efficient dispatch.

Our results for the horizontal auction can be viewed as an extension of Bikhchandani's (1995). He concludes that when several heterogeneous objects are sold simultaneously, with one auction for each type of object, every pure-strategy Nash Equilibrium is efficient. By creating an auction for each type of demand load, we have effectively recreated Bikhchandani's environment for an electricity auction, and extended upon his analysis by establishing the efficiency of the auction in a sequential setting.

4 Implementing Horizontal Auctions

The above analysis, although based on highly stylized models of the demand, generating cost and auction structure exposes a basic shortcoming of vertical auctions and suggest that horizontal auctions may be more compatible with the notion of self-commitment. In this section we will outline how a horizontal or a "Load Slice" auction may work in a real world environment. The description is intentionally vague recognizing the available flexibility in designing such auctions. We will first assume as in our stylized model a central market which ignores the locational characteristics of supply and demand and a fixed load curve which excludes demand side bidding. Later we will discuss how these assumptions could be relaxed. Under these restrictions the process may proceed as follows:

The auctioneer or exchange operator posts a load forecast for the bid period (say 24 periods) before the bidding process begins (few periods prior to the onset of the bid period). The auction is then done in several rounds filling up the load curve from the bottom up as illustrated in Figure 13. In each round the auctioneers solicits bids for a load slice of a specified number of GW with a fixed schedule prescribed by the demand curve (the initial solicitation is for base load dispatched for the entire bid period then for shoulder load and finally for peak load.) In each round bidders bid tenders consisting of capacity increments in GW that they are willing to commit to the specified schedule and a total price (or an average price per GWh). Since the schedule is known, bidders can easily calculate their total cost for serving a slice including all start-up costs and the costs associated with intertemporal constraints which they can factor into their bid price.

Winning bids in each round are selected based on lowest average price per GWh. Payments to the winning bids can be either discriminatory (i.e., paying each bidder their bid price) or uniform (i.e., paying all winning bids in a round the highest winning price or the lowest rejected price in that round). In either case it should be emphasized that the price per GWh will vary (most likely increase) as we move from lower base load slices to the upper peak load slices. The number of slices (and rounds) is a design parameter of the auction. The slices should be "thin" enough so that the dispatch schedule in each slice is approximately uniform.[11] This approach also allows different load slices to be auctioned at different time intervals; for instance, base load could be auctioned weekly while peak load could be auctioned twice a day. The appropriate time intervals for each load slice would be another design consideration.

The above design may be refined by allowing bidders to also specify a marginal energy price for minor adjustments to their energy output within a specified range (that would not affect the intertemporal costs). Such information might be useful in defining spot energy prices. As much as it is natural for the suppliers to define their tenders as horizontal slices of loads with prescribed schedules, it is natural for consumers to

[11] An alternative implementation could take the form of a Dutch Auction where the auctioneer posts a price per MWh for a load slice with a specified schedule and raises the price until a dispatch commitment is made. The auctioneer then moves onto the next slice, using the last accepted price as his starting bid, and continues the process until the entire load curve is filled.

think of electricity as a time differentiated commodity offered each period at a uniform spot price. Since energy consumed in a particular period cannot be traced to a specific supply source it should be sold at a uniform price. Economic efficiency dictates that the spot price at any period should be set at the marginal cost of the highest load slice active at that period.

Hence the market can be organized so that on the supply side power is acquired through a load slice sequential auction while on the demand side the energy is sold in a spot market with vertical slices priced uniformly in each time period (see Figure 13). It is possible to show that under some restrictive assumptions about the cost structure on the supply side and about the load pattern such a scheme will break even in the sense that the spot market sales will generate just enough revenue to recover the payments to suppliers (see Appendix). In general, however, the spot market will run a deficit (due to the intertemporal costs rolled into the supply bids) which must be recovered through a fixed charge or an "uplift charge."

The above scheme can be extended to include demand side bidding and locational factors. Both of these aspect will play a role during the peak load periods and need to be accounted for in the auction rounds for the upper slices. With regard to demand side bidding such bids can be entered as a reservation price in the upper auction rounds. In other words, if bid prices exceed the demand side bids all bids are rejected in favor of demand curtailments. The locational factors, i.e., congestion affects, can be accounted for by running a power flow analysis as the load curve is being filled up. When congestion occurs a locational penalty can be imposed as an adder to the bid price of upper slice bids originating at the constrained locations. This approach discriminates against peaking units in pricing congestion. Since the locational penalties are not imposed while the base load slices are being auctioned off, this amounts to giving priority to such units in congestion management. Intuitively it seems the right thing to do. However we have not yet analyzed the efficiency implications of such an approach.

5 Conclusions

Simulation studies and empirical evidence suggest that central unit commitment is inappropriate in a competitive electricity environment. Yet the traditional approach of auctioning power in an energy only hourly auction is incompatible with self-commitment due to intertemporal costs and nonconvexities in the production function for energy. One approach that has been adopted by the California Power Exchange proposal is a multi-round auction with activity rules which allow bidders to revise their bids so as to account for intertemporal costs resulting from their dispatch schedule. We propose a different approach which structures the auction so that the bid tenders allow bidders to readily account for these costs. Our proposed approach is supported by a game theoretic analysis of a stylized model which shows that unlike the vertical slice auctions, a sequential horizontal slice auctions will induce efficient dispatch.

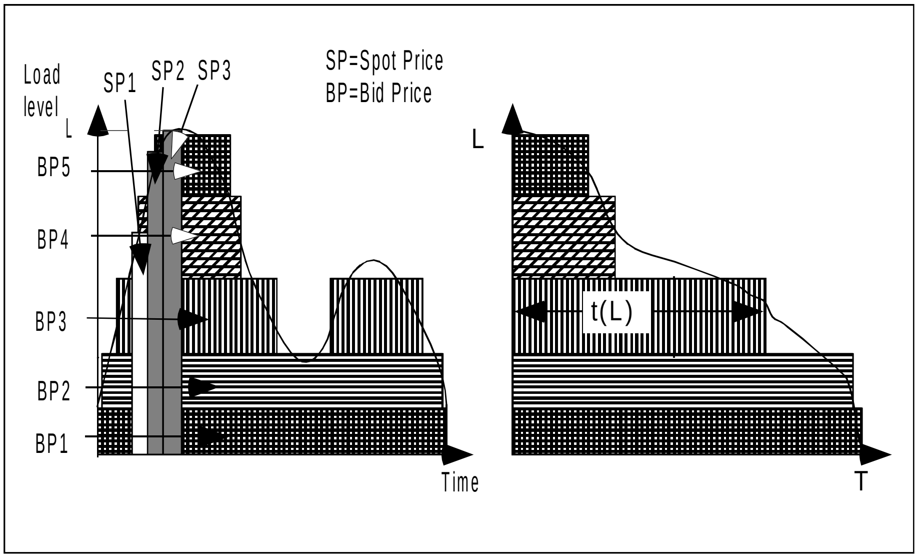

Figure 13: Load slice bidding with spot market selling

Appendix

The scheme mentioned is section 4 can be visualized as purchasing power to fill the load curve in uniformly priced "horizontal slices" were the price increases as the slice is higher in the stack. The power is then resold as uniformly priced vertical slices corresponding to different time periods, where the price for each slice is the spot price or the highest marginal price of the operating generation units.

We further assume that the load pattern is unimodal which implies at most one start-up in each load slice, and further assume that there is sufficient competition in each generation technology to make the optimal generation mix feasible. We ignore on/off switching aspects such as ramping-up and assume that a generator's total generation costs at a given load level depend strictly on the dispatch duration (this includes as a special case a two-part cost structure consisting of a start-up and marginal operating cost). Define the cost to generator type i of generating at capacity for t periods to be $C_i(t)$. The marginal cost of a generator may depend on the dispatch duration. We also assume that generation capacity of each type can be procured in any quantity which is a continuous variable.

In an idealized competitive load slice bidding environment, a generator will commit its capacity at a price per GWh equal to its average cost for the posted dispatch duration. Thus, each load slice will be served by the generating unit with the lowest average cost. Hence, if $t(L)$ denotes the duration corresponding to load level L, $\mathbf{L}$ is the maximum load level, T is the maximum duration, i.e., the base load duration, and the total cost of serving the load curve is given by

$$\begin{aligned} \text{Total Cost} &= \int_0^{\mathbf{L}} (\min_i \frac{C_i(t(L))}{t(L)})t(L)dL = \int_0^{\mathbf{L}} \min_i C_i(t(L))dL \\ &= \int_0^{\mathbf{L}} \widetilde{C}(t(L))dL = \widetilde{C}(T)L(T) + \int_{L(T)}^{\mathbf{L}} \widetilde{C}(t(L))dL \end{aligned}$$

where $\widetilde{C}(t) = \min\limits_i C_i(t)$ and $L(T)$ is the base load level. Splitting the integral in the last expression accounts for the fact that the function $L(t)$ is discontinuous at $t = T$ $\{L(T^+) = 0\}$.

Integrating by parts gives us

$$\text{Total Cost} = \widetilde{C}(T)L(T) + \widetilde{C}(t)L(t)|_T^0 + \int_0^{\mathbf{T}} \widetilde{C}'(t)L(t)dt = \widetilde{C}(0)\mathbf{L} \ + \int_0^{T} \widetilde{C}'(t)L(t)dt$$

If we include curtailment as one of the supply technologies then the function $\widetilde{C}(0) = 0$, so the first term in the above expression vanishes. Also $\widetilde{C}'(t)$ is the marginal cost of serving a load level whose duration is t. Since marginal cost is decreasing with duration, this is the highest marginal cost of generating units operating when the load level has duration t. Hence $\widetilde{C}'(t)$ is the spot price when the load level corresponds to duration t, and the integral of the spot price times the load over the entire time period equal the total cost. It thus follows that charging the spot price for the entire load in each time slice recovers total acquisition costs.

It should be noted that ramp-up costs and other costs associated with on/off switching cannot be recovered by spot prices that equal marginal costs. This is true in any system whether the acquisition is done through time slice bidding or load slice bidding, as described above. Recovery of switching costs requires some sort of "uplift" of the spot prices. In the above system generators are aware of these costs at the time bids are made and are able to internalize these costs. However, the spot prices will not recover the entire acquisition cots and some sort of fixed charge or uplift of the spot prices will be needed to achieve revenue neutrality.

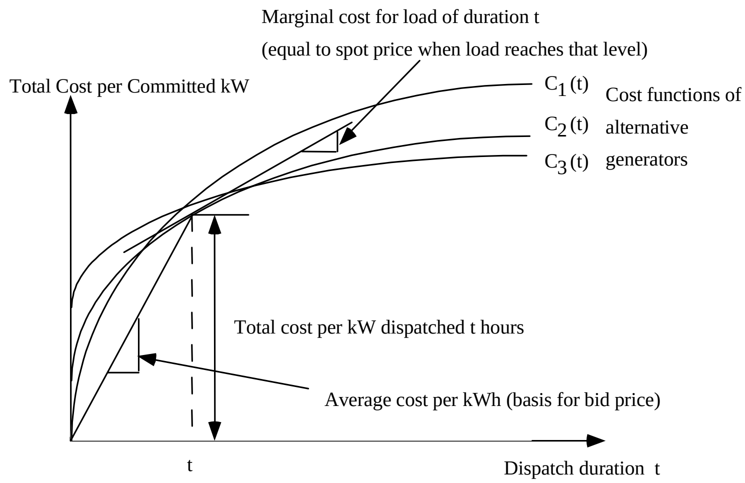

Figure 14: Optimal technology mix (given by the lower envelope).

References

Ausubel, L., and P. Cramton, 1996, "Demand Reduction and Inefficiency in Multi-unit Auctions," Working Paper No. 96-07, University of Maryland

Bikhchandani, S., 1996, "Auctions of Heterogeneous Objects," mimeo, The Anderson School, University of California at Los Angeles

Cramton, P., 1995, "Money Out of Thin Air: The Nationwide Narrowband PCS Auction," *Journal of Economic& Management Strategy,* Vol. 4, No. 2, pp. 267-343

Green, R. and D. Newbery, 1992, "Competition in the British Electricity Spot Market," *Journal of Political Economy*, Vol.100, No.5, pp.929-953

Hausch, D., 1986, "Multi-Object Auctions: Sequential vs. Simultaneous Sales," *Management Science*, Vol.32, No.12, pp.1599-1612

Isaac, R., D. James, 1998, "Robustness of the Incentive Compatible Combinatorial Auction," Maryland Auction Conference, May 29-31, 1998

Johnson, R., Oren, S. S., Alva J. Svoboda (1997). "Equity and Efficiency of Unit Commitment in Competitive Electricity Markets." *Utilities Policy*. Vol. 6, No. 1, pp. 9-19.

Maskin, E., and J. Riley, 1989, "Optimal Multi-unit Auctions," in Frank Hahn (ed) The Economics of Missing Markets, Information, and Games, Oxford: Oxford University Press, pp. 312-335

McAfee, P., and J. McMillan, 1996, "Analyzing the Airwaves Auction," *Journal of Economic Perspectives*, Vol. 10, No. 1, pp. 159-175

McMillan, J., 1994, "Selling Spectrum Rights," *Journal of Economic Perspectives,* Vol. 8, No. 3, pp. 145-162

Milgrom, P., 1998, "Putting Auction Theory to Work: The Simultaneous Ascending Auction," Maryland Auction Conference May 29-31, 1998

Patrick, R. H., Wolak, F. A. (1997). "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market." POWER Working Paper, PWP-047, University of California at Berkeley.

Rothkopf, M., A. Pekec, and R. Harstad, 1998, "Computationally Manageable Combinatorial Auctions," *Management Science*, Vol. 44, No. 8, pp.1131-1147

von der Fehr, N. and D. Harbord, 1993, "Spot Market Competition in the UK Electricity Industry," *The Economic Journal,* 103, pp. 531-546

Wilson, R., 1979, "Auctions of Shares," *The Quarterly Journal of Economics,* Nov., pp. 675-689

Wilson, R. 1998, " Efficiency Considerations in Designing Electricity Markets," Report to the Competition Bureau of Industry Canada