Extending the bidding format to promote demand response

Yanchao Liu a,*, Jesse T. Holzer b, Michael C. Ferris c

a Industrial and Systems Engineering Department, University of Wisconsin, Madison 53706, WI, United States
b Mathematics Department, University of Wisconsin, Madison 53706, WI, United States
c Computer Sciences Department, University of Wisconsin, Madison 53706, WI, United States

HIGHLIGHTS

- Three new bid types are proposed to enrich demand-side participation.
- Time value of electricity demand can be clearly conveyed to central dispatcher.
- The extended format preserves market efficiency and incentive compatibility.
- Energy storage is most effective to neutralize price volatility, with a limitation.

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form
15 June 2015
Accepted 16 June 2015

Keywords:
Demand response
Market design
Bidding
Economic efficiency
Energy storage
Convex optimization

ABSTRACT

We propose an extended bidding structure to allow more realistic demand characteristics and behaviors to be expressed via flexible bids. In today’s ISO-run energy markets, demand bid formats are all separable over time. However, a significant and growing segment of demand can be shifted across time and therefore has no way to bid its true valuation of consumption. We propose additional bid types that allow deferrable, adjustable and storage-type loads to better express their value, and thus elicit demand response in the most natural way – via direct participation in the market. We show that the additional bid types are easily incorporated into the existing market with no technological barrier and that they preserve the market’s efficiency and incentive-compatibility properties. Using real market data, we give a numerical demonstration that the extended bid format could substantially increase social welfare, and also present additional insight on storage expansion scenarios.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Sufficient demand-side participation is critical to the success of deregulated markets, since the marginal pricing and social welfare maximizing principles underlying their design are predicated on bid-based, competitive participation of both suppliers and consumers (Wellinghoff and Morenoff, 2007). However, reality has shown that the demand side lacks the ability to participate in the market comparably to the supply side, and exhibits significant unexpressed elasticity, resulting in inefficient market outcomes, exacerbating oligopoly power, and distorting long term investment incentives. There are two main causes. First, not all consumers are able to independently value the electricity ex ante (before the market clearing price is known) so as to place meaningful price-quantity bids on the market (Kirschen, 2003). This is inherent to the nature of electric energy, as most people regard electricity as an essential and non-substitutable commodity. Second, the bidding system does not provide a mechanism as an alternative to the price-quantity bid format for consumers to express their willingness to adjust consumption, particularly in response to price signals. Demand response (DR) is when a consumer modifies her usage behavior to account for price variations. For instance, if the consumer knows a priori that the price is high in some hours of the day and low in other hours of the day, she could reschedule usage to minimize the total cost (Schweppe et al., 1988). Incorporating changes in the market rules to induce demand response and encourage demand-side participation has garnered much recent attention from policy makers, practitioners and researchers.

Demand response resources are treated similarly to a generation resource by many ISO/RTO’s programs. For example, DR providers can specify operating requirements such as a minimum...
curtailment period and DR initialization cost. Energy bids are taken on a similar basis. Almost all ISO/RTOs in north America take demand-side energy bids exclusively in two forms\(^1\): (1) fixed, specified by a quantity in MWh, and (2) price-sensitive (or elastic), specified by a number of price-quantity pairs. These bids make the consumer act as a price-taker, or force her to provide an explicit demand curve. Kirschen (2003) notes that a normal consumer, and subsequently her wholesale market representative, e.g. load serving entity (LSE), are unable to estimate such curves accurately.

This paper proposes an extended bidding structure to encourage more demand-side market participation. The extended format enriches the forms of demand-side participation, promotes a broader frontier for load dispatchability and yet preserves the theoretical properties of the current market design philosophy, such as economic efficiency and incentive compatibility (Stoft, 2002).

1.1. Market model: theory and reality

While the specific formats proposed in this paper focus on the demand side, the structure can be applied to both sides of the market (e.g. a hydro generator may have time-shiftable supply needs). In the abstract form, each market participant \(k\) has a benefit function \(f_k(x_k)\) and operating constraint \(x_k \in X_k\), where \(x_k = [x_{k1}, ..., x_{kt}]^T\) is the energy consumption/supply schedule. The participant’s optimal response to the market price \(p\) is

\[
\max_{x \in X_k} f_k(x_k) - x_k^T p \tag{1}
\]

The solution \(x_k(p) = [X_{k1}(p), ..., X_{kt}(p)]\) determines the schedule of supply or demand across all times \(t = 1, ..., T\). Note that time dimension is embedded in the vectors \(x_k\) and \(p\), so all kinds of intertemporal relations can be expressed in the objective function as well as in the constraint \(x_k\).

In the bid-based central dispatch mechanism, each participant \(k\) simply informs (via bidding) the dispatcher its \(f_k(\cdot)\) and \(X_k\), and the central auctioneer (ISO/RTO) maximizes the social welfare by solving

\[
\max_x \sum_k f_k(x_k) \tag{2}
\]

s. t. \[\sum_k x_k = 0 \quad (\text{LP})\]  
\[x_k \in X_k, \quad \forall k \tag{3}\]

and using this model to set the market clearing prices. Eq. (3) says that the net power surplus (generation minus consumption) is zero and the market price \(p\) is the shadow price corresponding to this power balance equation. Note that each market participant has their own optimal response to these prices. If a dispatch and pricing model is designed such that the central dispatch solution with the accompanying prices coincides with the market participants’ optimal response to these prices, then competitive participants have every reason to bid their true parameters,\(^2\) thus the model is incentive compatible. If the central dispatcher does not have accurate input \(f_k(\cdot)\) and \(X_k\) from market participants about their true valuation of electricity, there is no way for the dispatcher to maximize the social welfare. In other words, one cannot maximize something without accurately measuring it.

The existing market model (where only fixed and elastic bids are allowed) is a special case of (1), having two specialities:

1. the value function \(f\) is separable across time, thus restricted to the form

\[f_k(x_k) = \sum_{t \in T} f_{k,t}(x_{k,t})\]

2. the constraint set \(X_k\) of a consumer \(k\) is also separable across time, i.e.,

\[X_k = \prod_{t \in T} X_{k,t}\]

These restrictions hinder efficient market participation. For example, a consumer with shiftable demand has no way to express this shiftability in a bid and may have to predict the price path so as to approximate this feature using time-separable price-quantity bids. The prediction and approximation are error-prone and most likely to lead to suboptimal outcomes.

In contrast, the general (and well-established) model above avoids such problems but still retains nice theoretical properties. As long as each \(f_k(\cdot)\) is a concave function, and each \(X_k\) is a convex set, the important economic design properties will hold and the model will remain easy to solve. The key point of this paper is to propose new mechanisms for bidding that allow for more complex \(f_k\) and \(X_k\) to be described in natural ways. In addition to fixed and elastic demand, we identify three additional types of demand, namely, shiftable, adjustable and arbitrage. We will formulate the basic characteristics and model the behavior of each type of demand. Fig. 1 illustrates a structural overview of our work. Note that all of our formulations are in the form given by (1) for particular choices of \(x_k\), \(X_k\) and \(f_k\), and the new types are not separable across time. This amounts to a policy change that enhances the types of load that can be bid into the market.

1.2. Policy and literature review

Broadly, a more responsive demand-side is desired from both economic and reliability standpoints. However, the existing policy on demand response (FERC, 2011), which requires ISO/RTOs to compensate curtailed energy consumption at the LMP, sends wrong economic incentives to market participants.\(^3\) A full-LMP compensation gives the DR providers, presumably retail customers who would normally be charged at the retail price \(G\) for consumption, both the LMP and the savings from not consuming, totaling \(LMP + G\), which amounts to an uneconomic double-payment, see Chao (2010, 2011) and Hogan (2009, 2010, 2012). Within the monetary compensation framework, ISO/RTOs have made various localization efforts to preserve the economic efficiency of demand response. For example, ERCOT has implemented an “LMP minus proxy G” approach to avoid the double-payment problem, where \(G\) is a proxy for the purchase price or contract price that is
This is unlikely to be true and accurate in general, representative of what retail customers paid for their energy adjusted for risk. PJM RTO treats the emergency demand response as a reliability resource where curtailment must strictly follow the RTO’s dispatch orders. For voluntary load reduction, compensation is only assessed when the wholesale price is higher than a net benefit price published monthly by the RTO.

Compared to the monetary compensation framework, improving the bidding format has shown great potential for promoting demand-side participation. Arroyo and Conejo (2002) present a unit commitment based market clearing mechanism that is widely used in today’s markets. The mechanism encourages consumers to submit price-quantity bids to the market operator, instead of being treated as fixed and rigid. Strbac and Kirschen (1999) demonstrate the importance of a realistic demand-side bidding structure. They stress that the cost of load recovery after, or occasionally before, the load reduction period should be accounted for in an optimal schedule. Su and Kirschen (2009) propose a complex form of demand bids that allowed for flexible time of consumption. In particular, consumers could submit multiple price-quantity bids for each consumption period, and specify the total amount of consumption to be satisfied over the scheduling horizon. Papadaskalopoulos et al. (2011) present a decentralized market clearing mechanism in which each market participant computes her own optimal generation or consumption schedule and bids given the market prices. The central planner in turn updates the prices based on the bids from market participants. This is an iterative process and the iteration proceeds until an equilibrium is reached. We recognize the merits of the freedom to interpret and respond to price signals for market participants, but such freedom may render an equilibrium nonexistent and/or the iterative process may not converge. We believe that a certain degree of conformity is as important as flexibility in the design of a bidding structure, and we postulate that adding new bidding formats to existing practice is a more controlled and easier way to implement change than going to an iterative process.

The remainder of this paper is organized as follows. Section 2.1 proposes our characterization of different demand types and their respective cost-minimizing or surplus-maximizing properties. Based on this, Section 2.2 develops the new bidding structure and the corresponding central dispatch model, accompanied by the discussion of its incentive compatibility. Section 3 implements the model for a series of experiments and presents the experiment results. In Section 4, we discuss additional technical points related to the implementation of the model in practice, including network integration, unit commitment and other practical issues. Section 5 concludes the paper and summarizes the policy implications of our work.

2. Methods
Symbols will be defined where they first appear in the paper. In general, \( g \) and \( d \) denote generation and demand in megawatt hour (MWh), respectively, and \( p \) denotes the price in dollars/MWh. The superscript on a symbol annotates a specific meaning, while subscripts \( k \) and \( t \) index the participant and time period (i.e. hour), respectively. Depending on the context of its occurrence, a symbol may represent a scalar or a vector, with the specific meaning implied by the presence or absence of the subscripts. A symbol topped with a bar or underlined is always a parameter instead of a variable, representing the upper or lower bound of a quantity.

2.1. Demand types and behavioral models

In this section, we outline existing and new types of demand bids and show how they have a natural parameterization for the user while fitting nicely into the general framework outlined in Section 1.1. The new bid types allow a bidder to express a richer variety of these choices, and hence enable the market to correctly value these possibilities.

2.1.1. Fixed demand
Fixed demand constitutes a dominant portion of the total demand on the spot market. For example, in MISO’s day-ahead market in 2008, fixed demand bids accounted for about 98% of total cleared demand (Newell and Hajos, 2010). By submitting a quantity without putting a maximum acceptable price, the bidder effectively tells the market that she places an infinite value on the specified amount of electric energy, i.e., \( x_k = \{d_k\} \). This is unlikely to be true and accurate in such an overwhelming scale, but it is what is happening on the market every day.

Using fixed demand bids in cases where additional flexibility is present unnecessarily restricts efficiency and should be discouraged. Fixed demand bidders have nothing to offer.
optimize because they are unconcerned about the price.

2.1.2. Elastic demand

Elastic demand, unlike fixed demand which registers a straight vertical line, exhibits a monotonically non-increasing curve on the price-quantity plot. The value (or utility or benefit) is a concave function (non-increasing marginal value) of the consumption \( d \), denoted by \( V(d) \). Note that the value function can be different for different time periods, but it is separable with respect to the time of consumption. The surplus maximization problem of a consumer \( k \) with elastic demand is (ELA) (\( p \)):

\[
\max_{d_k} \sum_t [V_k(d_{kt}) - p_t d_{kt}]
\]

s. t. \( d_{kt} \leq d_{kt}^+ \leq \bar{d}_{kt}, \quad \forall t \) \eqno (5)

Typical forms of \( V(d) \), like those of the generator cost function \( C(g) \), are quadratic or piecewise linear. The associated bid is \( X_k = \{ d_k, \bar{d}_k \} \) and \( p_k(x_k) = \sum_t V_k(x_{kt}) \).

2.1.3. Adjustable demand

Similar to fixed demand, a consumer with adjustable demand has a preferred consumption profile, but is willing to make adjustments at a cost. For example, shopping malls usually have established standards for the level of lighting and HVAC, but the inconvenience cost of deviating from the standards can be estimated as well. Let \( r_{kt}^+ \) and \( r_{kt}^- \) denote the amount of over- (adjust up) and under- (adjust down) consumption from the target level \( x_{kt} \), respectively, and let \( D_{kt}(r_{kt}^+, r_{kt}^-) \) denote the deviation cost. Over-consumption does not normally incur extra costs on the consumer's side, and we include its cost here simply for the generality of the formulation since it does not lead routinely to extra benefits. Compared to the value function of elastic demand, the deviation cost function is an alternative valuation of electric energy, also termed as the Value of Lost Load (VOLL)\(^6\). An arbitrageur seeks to profit from the price discrepancies over time – buy energy when the price is low, store it, and sell when the price is high. There are no target levels of storage and no deviation penalties, but there is efficiency loss in the charge-discharge cycles. Let \( s_{kt} \) and \( b_{kt} \) denote sell (discharge) and buy (charge), respectively, and \( h_{kt} \) denote the storage level. An arbitrageur maximizes its profit by solving (ARB) (\( p \)):

\[
\max_{b_{kt}, s_{kt}, h_{kt}} \sum_t p_t (s_{kt} - b_{kt}) \]

s. t. \( h_{kt-1} = h_{kt-1} + b_{kt} - s_{kt}, \quad \forall t \) \eqno (15)

\[
h_{k0} = h_{k1}, \quad 0 \leq s_{kt} \leq s_k, \quad \forall t \quad \text{(16)}
\]

\[
0 \leq b_{kt} \leq b_k, \quad \forall t \quad \text{(17)}
\]

\[
0 \leq h_{kt} \leq h_k, \quad \forall t \quad \text{(18)}
\]

In the defining equation (15) for \( h_{kt} \), \( e_k \) is the efficiency factor with \( e_k \in [0, 1] \), indicating that each unit of energy input will convert to \( e_k \) unit of output. Realistically, \( e_k \) may be a function of \( h_k \), e.g., the efficiency of a Sodium Sulfur (NaS) battery depends on the depth of discharge (Himelic, 2011), which needs more constraints to express. For expositional purpose, we make \( e_k \) a constant bidding parameter, but a non-constant \( e_k \) may have implications on the applicability of this model. Constraint (16) fixes the net change of the storage level \( h_k \) over the planning period (e.g., a 24-h period for the day-ahead market) to zero for sustainable operations, although in practice this constraint can appear in different forms.

Linking back to (1), \( X_k = \{ b_k, s_k \} \), \( X_k = \{ d_{kt}^-, \bar{d}_{kt} \} = b_{kt} - s_{kt} \), with (15) – (19) which is a polyhedral convex set and \( f_k(\cdot) \equiv 0 \). To bid electricity to be delivered within a given time range, and is flexible with regard to the time of delivery within that range. For instance, consumer \( k \) partitions the planning horizon \( T \) into time ranges indexed by \( m \), and requires the amount \( d_{kt}^m \) to be delivered within the time range \( t \) \( \in T \). A consumer with shiftable load minimizes her consumption cost by solving (SHI) (\( p \)):

\[
\min_{d_{kt}^m} \sum_t p_t d_{kt}^m \quad \text{s. t. } \sum_t d_{kt}^m = d_{kt}^m, \quad \forall m, \quad t \in T_m \]

\[
d_{kt}^m \leq d_{kt}^m, \quad \forall t \quad \text{(13)}
\]

The shiftable demand bid requires no explicit valuation of the electricity, and opens a door for consumers to respond to the market prices. Such demand can be expected to substitute for an appreciable portion of the fixed demand, and hence increase the general dispatchability of the demand. Typical shiftable loads include plug-in electric vehicles (PEV) and their aggregators, industrial laundry facilities and sewage treatment plants, etc. More examples, and an interesting mechanism to deal with deferrable load, can be found in Papavasiliiou and Oren (2008).

Linking back to formulation (1), \( X_k = \{ d_{kt}^m \} \) (12) and (13) hold } and \( f_k(\cdot) \equiv 0 \). The bid is parameterized by \( T_m, d^u, d^b, d^sh \).
by adjusting the prices $p_t$.

We postulate a central dispatch model of the form (2)–(4), as follows.

Central Model:

$$\min_{g,k,d,h,t,s} \sum_{k,t} \left[ C_{k,t}(g_{k,t}) - V_{k,t}(d_{k,t}) + D_{k,t}(b_{k,t}) \right]$$

subject to

\begin{align*}
\text{(6),(12),(13),(8),(9),(15)} & - (19) & - (21) & - (24). \\
\end{align*}

The price $p_t$ is set as the optimal Lagrangian multiplier (or dual variable) of the corresponding constraint in (24). Because the model takes the true cost/benefit functions as input and minimizes the total social cost, it is economically efficient.  

**Theorem 1.** Given a set of bidding parameters, suppose that

$$\hat{x} = (\hat{g}, \hat{d}, \hat{a}^s, \hat{a}^h, \hat{p}^s, \hat{p}^h, \hat{b}, \hat{s}, \hat{h})$$
solves the Central Model and $\hat{p}$ is the optimal Lagrangian multiplier of the constraint (24). Then $\hat{x}$ solves (GEN)$\hat{p}$, $\hat{d}$ solves (ELA)$\hat{p}$, $\hat{a}^s$ solves (SHI)$\hat{p}$, $(\hat{p}^s, \hat{p}^h)$ solves (ADJ)$\hat{p}$, and $(\hat{b}, \hat{s}, \hat{h})$ solves (ARB)$\hat{p}$.

The theorem is a straightforward result from duality theory so we forgo the proof. It says that the Central Model will give the same dispatch as the one obtained from individual market participants maximizing their self-interest. It is widely believed that this property of the economic dispatch model, coupled with the reality that non-convex cost (e.g., unit commitment cost) is relatively minor, makes the existing bidding structure incentive compatible, see Stoft (2002) and Groves and Ledyard (1985) for a survey of the notion of incentive compatibility. This leads to the conclusion that the extended Central Model is incentive compatible.

### 3. Results

While the proposed model opens up new ways for demand bidding, the actual penetration rate of the new demand forms is yet to be seen, and the exact bidding parameters are still unknown. In the experiments, these bidding parameters are determined so as to (1) fit sensibly within the base-case data set, (2) agree with the modeling features and (3) approximate the practical reality as closely as possible. Detailed explanations are given below.

#### 3.1. Data and setting

Under the existing two-settlement scheme of today’s ISO-run markets, the target market of application is the day-ahead market (instead of the real-time market) where the temporal dimension $t$ represents hours. Therefore, we will use data from a day-ahead market in the experiments. The generator bids and the fixed demands are obtained from the FERC elibrary Docket Number AD10-12, ACCNUM 20120222-4012. The data set represents a typical summer operating day of the PJM day-ahead market (Krall et al., 2012). For the demand data, we sum up the fixed demand bids from all 13,760 buses for each hour to create an aggregate hourly demand profile, for use as the base case in the experiments.  

The base case is illustrated in Fig. 2 as the “Fixed”
demand. For the generator data, there are altogether 1011 generators, each offering up to 10 pairs of price-quantity bids for energy, and various unit commitment requirements and costs. A unit commitment process similar to the one documented in Krall et al. (2012) was executed on the base-case demand, which selected 365 generators for commitment. We fix the unit commitment status according to this result in the subsequent experiments.

We make up four aggregate consumers, one for each demand type. The omission of subscript $k$ in the following should cause no confusion.

3.1.1. Elastic demand

We assume that 1% of each hour’s base-case demand becomes elastic, which is then bid into the market in ten equally sized MWh blocks, coupled respectively with 10 decreasing prices ranging from $99/MWh to $0/MWh with even decrements, see Fig. 3 for an illustration. This piece-wise linear demand curve for hour $t$ is represented by a linear cost function $V_t(d_{t})$ and two linear constraints in the minimization problem, as follows:

$$V_t(d_t) = \sum_{o \in O} p_{t,o}^{db} d_{t,o}^{db}$$  \hspace{1cm} (25)

$$d_t = \sum_{o \in O} d_{t,o}^{db}$$  \hspace{1cm} (26)

$$d_{t,o}^{db} \leq \bar{d}_{t,o}^{db} \quad \forall \ o \in O$$  \hspace{1cm} (27)

where $O$ is the set of bid blocks, the bidding pair $(p_{t,o}^{db}, \bar{d}_{t,o}^{db})$ indicates that an increment of $\bar{d}_{t,o}^{db}$ MWh is worth $p_{t,o}^{db}$ dollars/MWh to the consumer, and the variable $d_{t,o}^{db}$ represents the dispatched quantity in bid block $o$.

![Fig. 2. Day-ahead demand profile.](image)

![Fig. 3. Elastic demand bid for hour 1.](image)
3.1.2. Adjustable demand

We assume 1% of each hour’s base-case demand becomes the target level $d^A_t$ of the adjustable demand. The deviation function $D_r(t^*_r, r^-)$ is taken in the form of (10), with the linear penalty $\beta^+_t$ and $\beta^-_t$ arbitrarily set to the minimum LMP (0) and the average LMP (30.1) of the base-case, respectively, and the quadratic penalty $\alpha^+_t$ and $\alpha^-_t$ arbitrarily set to 0.05 and 0.1, respectively. Note that the choice of these factors reflects the relative merits of over-consumption compared to under-consumption. The bound $r^-_t$ is set equal to $d^A_t$ while $r^+_t$ is set to $\sum d^A_t$.

3.1.3. Shiftable demand

We partition the 24-h period into three 8-h ranges, i.e., $T_m$, $m = 1, 2, 3$, and assume 1% penetration of shiftable demand by setting the total demand requirement $d^S_m$ for range $m$ to be 1% of the sum of the hourly base-case demand in the range.

3.1.4. Arbitrage

We assume an arbitrageur (storage) at the size of 1% of the base-case demand is present besides the base-case demand, and set $\bar{h}$ accordingly. We set the hourly buy (charging) rate $b$ and sell (discharging) rate $s$ to be $0.2\bar{h}$, to mimic the characteristics of a 5-h storage facility. The efficiency factor $e$ is set to 0.75.
3.2. Comparative effect of different demand types

We tested the effect on LMP and social welfare of 1% penetration of the outlined forms of demand-side bids. The elastic, shiftable and adjustable demands are substitutes for the fixed demand, so the fixed demand will reduce to 99% of the original level in these individual cases. The results are represented in the first four rows of Table 2. The table lists the cost (negative of the social welfare) results. The first column indicates the hypothesized market composition, the second column is the cost from the current bidding design, the third column is the optimal cost from our proposed bidding design, and the fourth and the fifth columns compare the costs, and list the savings and percent savings, respectively. Even for small penetrations, the proposed bidding design yields apparent and significant benefits.

Arbitrage is an additional form of participation on top of the base-case demand, so the base-case demand remains at the 100% level. We examined two aggregate cases, both consisting of 97% fixed demand and 1% each of the elastic, shiftable and adjustable demand, one with 1% arbitrage and the other without arbitrage. The final two rows of Table 2 show that the arbitrage bids generate even more significant savings. Note that additionally we plot the dispatched demand of the aggregate case (i.e., the “97% Fixed + 1% (E+S+A+AR)” case) as the “Dispatched” curve in Fig. 2. This shows that the new bidding flexibility reduces the variability in system load over time.

Figs. 4 and 5 plot the LMP resulting from different cases. As expected, the base case exhibits the roughest (with the biggest dip and spike) price path while the aggregate case exhibits the mildest. The penetration of each individual demand type smoothens

---

Fig. 6. LMP for different arbitrage levels.

Fig. 7. Profit of arbitrage for different penetration and efficiency.
the LMP to a certain extent. Among them, arbitrage is the most effective, followed by shiftable demand. Elastic demand is the least effective in terms of dampening the price fluctuation.

3.3. Arbitrage effect on the LMP and profit

As demonstrated above, arbitrage has the most impact on the LMP among all participant types at the same penetration level, i.e., 1%. This is understandable, as an arbitrageur’s buy/sell schedule is driven solely by the temporal price differences, and not by any target level of consumption or private valuation of the electric energy. However, unlike the other types of demand bids which are direct alternatives or substitutes for the fixed demand bid, the arbitrage bid must be backed by physical storage capability that takes time to construct and deploy, so the penetration level is likely to be small in the foreseeable future.

In Fig. 6, we plot the effect of arbitrage on the LMP for different penetration levels, ranging from 0.2% to 1%. As expected, the increase of the arbitrage level will dampen the LMP variation. It is also interesting to note that the effect does not grow linearly with the penetration level, e.g., the first 0.2% increment of arbitrage contributed about half of the peak price reduction. This observation prompts a question: what is the “optimal” percentage of storage on the market? Fig. 7 provides some useful information to address this question.

Fig. 7 plots the profits of arbitrage for penetration levels ranging from 0% to 2% with an increment of 0.1%, and for three different efficiency factors, i.e., 0.65, 0.75 and 0.85. As can be seen from the figure, high marginal value of storage expansion can be expected when the level is below 0.4% for all three efficiency options. From a level higher than 0.6%, the marginal benefit of expanding storage capacity can become negative, indicating that further expansion of storage capacity is uneconomical. Of course, in making the storage expansion decision, construction and operation costs and a myriad of other factors need to be considered, but the above observation at least sheds some light on such a decision-making process.

4. Discussion

The implications of our model on the market have been demonstrated in the previous section. However, there are some additional points that we wish to discuss regarding the implementation of the model in today’s market realities.

4.1. Network integration

The above framework is developed only on an economic basis, without the use of transmission network variables and constraints. This is purely for the clarity of the main point. In fact, the framework can be easily adapted to a DC-based (linearly constrained) network model, and the theoretical properties will continue to hold. Suppose the network is represented by a set of nodes $\mathcal{N}$ and a set of arcs $\mathcal{A}$ (each physical transmission line is modeled by two arcs, one for each direction). Let the variable $\delta$ denote the power flow on arcs, bounded within the thermal limits $[-2, 2]$, the variable $\theta$ denote the voltage angle at nodes, and the parameter $B$ denote the susceptance of arcs. Then the system operator maintains the arc flow equation and the nodal power balance, as follows:

$$\begin{align*}
\sum_{\delta(i, l)} - \left[ B_{k}(\delta_{k} - \delta_{kl}) \right] & = 0, \quad \forall (k, l) \in \mathcal{A}, t \\
\sum_{k} \delta_{kl} & = 0, \quad \forall k, t
\end{align*}$$

It is easy to see that these additional variables and linear equations can be readily incorporated in the Central Model (2)–(4) and connect readily to the variables used in (5)–(23).

4.2. Unit commitment

The general belief that the bid-based two-sided market design is economically more efficient than the vertically integrated utility model is based on certain assumptions, including that the market surplus can be accurately measured (e.g., one knows exactly at what price one would stop using electricity), that the market is competitive and that costs are convex, etc. However, practical market operations have to account for many real-world situations where those assumptions do not perfectly hold. In general terms, most ISO-run market operates like this: the ISO takes bids and offers, mitigates offers then maximizes the market surplus by running a unit commitment economic dispatch model. After determining the optimal commitment and dispatch, the ISO announces the LMPs and dispatch quantity. It guarantees make-whole payments to generators and announces the final settlement price, ex-post. We believe that the gap between theoretical design and practical operation is relatively minor and does not overshadow the merits of the bid-based market model.

Although the proposed demand bid structure is developed on the theoretical economic dispatch model, it can be applied to ISO’s operational models as well.

Using the notation of the abstract model (1), we now briefly describe the extension into unit commitment, where not only the level of consumption or production by each participant is determined, but also the decision of whether to consume or produce at all. The decision of participant $k$ to produce or consume (generally to be active) in time period $t$ is represented by the binary variable $u_{kt}$ ($=1$ if active). This logic is embodied in constraints $g_{k}u_{kt} \leq x_{k} \leq g_{k}u_{kt}$, where $x_{k} \in [x_{k1}, x_{k2}]$. More complex constraints on $u$, such as minimum up-time, are represented abstractly by $u_{k} \in U_{k}$. Finally the benefit of activity to participant $k$ over the planning horizon is modeled by $g_{k}(u_{k})$. This function might include, for example, startup and shutdown costs. Then the complete unit commitment problem is

$$\begin{align*}
\text{max}_{x_{k}} & \quad \sum_{k} f_{k}(x_{k}) + \sum_{k} g_{k}(u_{k}) \\
\text{s. t.} & \quad \sum_{k} x_{k} = 0 \ (\text{LP}) \\
& \quad g_{k}u_{kt} \leq x_{k} \leq g_{k}u_{kt} \\
& \quad x_{k} \in X_{k}, \quad u_{k} \in U_{k}
\end{align*}$$

Note that consumers as well as suppliers are allowed to bid unit commitment characteristics $U_{k}$ and $g_{k}$. Even without load shifting and other complex bids by consumers, the unit commitment problem poses a number of theoretical difficulties, for example, the lack of an equilibrium price as well as a canonical payment rule (Johnson et al., 1997; O’Neill et al., 2005; Liu and Ferris, 2013). The prices $p$ are defined as the Lagrangian multipliers on the power balance constraints obtained by fixing the commitment $u$ to an optimal value and solving the remaining convex optimization problem (O’Neill et al., 2005), but these prices alone do not support a socially optimal commitment in market participants’ utility maximization models. Indeed, no uniform price has such a property, and uplift payments are necessary to secure sufficient participation by suppliers. Relative to the overall market size and complexity, the theoretical difficulties are inconsequential and have been effectively neutralized in ISO’s market rules. We believe that allowing consumers to bid into unit commitment will
not exacerbate the problem, and the practical solutions currently in use for suppliers will continue to be adequate even if they are extended to consumers.

4.3. Other practical realities

We acknowledge that the proposed extension in bidding structure cannot eliminate the intrinsic difficulty for residential consumers to value electricity as a commodity. Nonetheless, the extension could enrich LSEs’ options in acquiring bulk power through the wholesale market. This flexibility, coupled with increasing competition (e.g., competition with distributed power systems such as rooftop solar panels and microgrids) in the retail market, could drive LSEs to innovate in their DR programs which would eventually steer (real-time) end-use behaviors toward the right direction.

Finally, we note that accepting bids from both sides of the market is not the only way grid-based energy markets can be efficiently organized. In ERCOT, for example, loads can “passively” respond to prices as price-takers. By doing so, they do not contribute to price formation in the dispatch algorithm and hence need not worry about penalties for deviating from any forecast or scheduled load levels. By exempting loads from any form of day-ahead commitment for consumption, the ISO assumes a higher level of responsibility in load forecasting while dispatching generation resources. From an optimistic perspective, if all consumers are exposed to the real-time prices and are free to respond, the long-run market outcome will settle at the efficient equilibrium. However, two market realities currently prevent the full realization of such a scenario: (1) the lack of advanced metering and control infrastructure, and (2) the use of fixed uniform retail rates in the regulated retail market (Chao, 2010). Therefore, there are numerous opportunities remaining for infrastructure innovation and policy reforms.

5. Conclusions and policy implications

A properly functioning market should reflect both the willingness of sellers to sell at a price and the willingness of buyers to purchase at a price. In an RTO- or ISO-run market, however, buyers are generally unable to directly express their willingness to pay for the electric energy at the price offered. RTOs and ISOs cannot isolate individual buyers’ willingness to pay which results in extremely inelastic demand. A simple monetary compensation rule designed to elicit demand curtailment, as prescribed in the controversial FERC Order 745, has not been successful in providing a satisfactory solution at this time. This is because demand response is not an inherent element of competitive energy markets; rather, it is a recourse measure to account for the market imperfection, which is partly attributed to the unnecessarily restrictive bidding structure.

We argue that an effective policy is to open up the bidding structure and allow a greater variety of realistic demand characteristics to be expressed. To this end, we proposed an extended bidding structure and the corresponding market clearing model as a measure to promote demand-side participation in wholesale energy markets. In particular, the model accounted for three new types of demand characteristics, namely, shiftable, adjustable and arbitrage. The outcome of the proposed mechanism was shown to be economically efficient and incentive compatible, both of which are critical design principles for ISO-run energy markets.

If the new mechanism were put to use, significant economic benefit could be realized even from a small market presence of the proposed new bid types. Among different bids, energy storage is most effective in dampening price volatility. However, the marginal effect of storage decreases as the installed capacity increases and a capacity of 0.6% of the total load is about the maximal economical level. These conclusions have been demonstrated in a simulated market clearing environment.

Compared to the vacated FERC policy which offered wrong economic incentives to market participants, our model provides a justifiable approach to promote demand-side participation. Furthermore, the proposed model does not require a drastic change to the existing auction-based market-clearing mechanism and poses no technological barrier for its practical deployment. In the meantime, we point out that the particular model presented in this paper is a prototype of a new design philosophy within the confines of an efficient market, while further extension and customization are both possible and practical, promoting much more flexible demand-side participation in the market.

Acknowledgments

This work is supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract number DE-AC02-06CH11357, and in part by Air Force Grant FA9550-10-1-0101. The authors would like to thank the editor and the three anonymous reviewers for their contribution for improving the quality of this paper.

References


