Economic Dispatch of Demand Response Balancing through Asymmetric Block Offers

<table>
<thead>
<tr>
<th>Journal:</th>
<th>IEEE Transactions on Power Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Transactions</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>O’Connell, Niamh; Technical University of Denmark, Informatics and Mathematical Modeling</td>
</tr>
<tr>
<td></td>
<td>Pinson, Pierre; Technical University of Denmark, Department of Electrical Engineering</td>
</tr>
<tr>
<td></td>
<td>Madsen, Henrik; Technical University of Denmark, Department of Applied Mathematics and Computer Science</td>
</tr>
<tr>
<td>Technical Topic Area:</td>
<td>Power system operations, Power system markets</td>
</tr>
<tr>
<td>Key Words:</td>
<td>Electricity Markets, Demand-Side Management, Mixed-Integer Linear Programming, Refrigeration, Time Series Analysis</td>
</tr>
</tbody>
</table>
Economic Dispatch of Demand Response Balancing through Asymmetric Block Offers

Niamh O'Connell, Student Member, IEEE, Pierre Pinson, Senior Member, IEEE,
Henrik Madsen, Member, IEEE, and Mark O'Malley, Fellow, IEEE

Abstract—This paper proposes a method of describing the load shifting ability of flexible electrical loads in a manner suitable for existing power system dispatch frameworks. The concept of an asymmetric block offer for flexible loads is introduced. This offer structure describes the ability of a flexible load to provide a response to the power system and the subsequent need to recover. The conventional system dispatch algorithm is altered to facilitate the dispatch of demand response units alongside generating units using the proposed offer structure. The value of demand response is assessed through case studies that dispatch flexible supermarket refrigeration loads for the provision of regulating power. The demand resource is described by a set of asymmetric blocks, and a set of four blocks offers is shown to offer cost savings for the procurement of regulating power in excess of 20%. For comparative purposes, the cost savings achievable with a fully observable and controllable demand response resource are evaluated, using a time series model of the refrigeration loads. The fully modelled resource offers greater savings, however the difference is small and potentially insufficient to justify the investment required to fully model and control individual flexible loads.

Index Terms—Mixed-Integer Linear Programming, Electricity Markets, Demand-Side Management, Refrigeration, Time-Series Analysis.

I. INTRODUCTION

Demand response is frequently presented as a solution to a multitude of challenges in the power system. It is said to bring about such benefits as supporting higher penetrations of renewable generation [1], increasing economic efficiency [2], and alleviating distribution network congestion [3], among others [4], [5]. Demand response is not without its challenges however. Chief amongst these is the uncertainty over the value that demand response provides to the power system.

A number of academic works have attempted to quantify this value. The concept of price elasticity of demand is often adopted as a representation of the flexibility of demand in the presence of dynamic or real-time prices [6]–[8]. This approach assumes economic rationality and overlooks the significant complexities of electrical demand. Demand response is fundamentally characterised by the physical limitations and dynamics of electrical end-uses and highly complex interaction with consumers, which are not accurately described in the form of a linear price curve or single elasticity value.

On the other end of the scale, detailed models are used to assess the abilities and value of demand response resources, where it is assumed that the internal states of individual resources are known and can be controlled [9]–[12]. This approach is a valid method of establishing the theoretical value of demand response, however, the financial and computational costs of establishing a framework to dispatch many thousands of individually modelled flexible loads are prohibitive. Furthermore, current market clearing and power system dispatch algorithms interface with large conventional generating resources through bids consisting of a volume and a price [13], or limited set of constraints [14]. These frameworks are unsuitable for the management of a large volume of individually modelled and controlled flexible loads.

This paper demonstrates the value that demand response can provide to the power system when its representation in the system dispatch algorithm is limited to one that is comparable in complexity to that of conventional generating units. The representation described here is suitable for the interface between an aggregator, managing a population of responsive loads, and the market or system operator. The interface between the aggregator and the individual loads can be handled using such control frameworks as detailed in [11], [12].

This work contains two novel contributions to the field of demand response research. Firstly, building on material presented in [15], we develop a methodology of defining block bids that populations of flexible demand units can offer to the power system or market operator. These block bids reflect the load shifting abilities and limitations of individual demand units, and consider both their ability to provide a response for either up- or down-regulation, and the need to recover following the response.

Secondly, we present an optimisation framework to dispatch these block offers for demand response alongside conventional generating units within an economic dispatch framework for the provision of regulating power. This optimisation differs from conventional economic dispatch algorithms as it is a combinatorial prob-
lem, where demand response blocks can be accepted in their entirety, or not at all.

Case studies are conducted to assess the value of demand response when represented by a limited set of block offers in the system dispatch algorithm. A comparative study evaluates the demand response resource when described as a fully observable and controllable system, using a time-series model. The flexible load considered in these case studies is supermarket refrigeration, which has with significant potential for load shifting demand response [16]–[18].

The remainder of this paper is structured as follows. Sections II and III present the demand response model, both the full and limited representations. The optimisation framework employed to dispatch the system considering demand response is detailed in Section IV. The case study framework is outlined in Section V and results are given in Section VI. Concluding remarks can be found in Section VII.

II. DEMAND RESPONSE RESOURCE MODEL

In this work we consider load shifting demand response on a short-term horizon, specifically for the provision of regulating power. A number of load types are considered as candidates for load shifting. In particular, thermal-electric loads such as building heating and cooling [10], [19], water heating [20], and refrigeration [16]–[18], are considered ideal candidates due to their ability to alter their power consumption while maintaining an acceptable temperature range. These thermal loads share two key characteristics, namely, saturation and rebound. Saturation refers to the limited time extent of the response from a thermal load. This is due to the temperature constraints that limit the duration for which power can be adjusted either upwards or downwards from a given baseline. Rebound refers to the phenomenon that is observed after control is returned to the device from the aggregator. Upon return of control, the device will attempt to return to the state it occupied directly preceding the request from the aggregator, resulting in a sudden return in the direction of power consumption.

Supermarket refrigeration is employed here as a representative flexible thermal-electrical demand. Supermarket refrigeration systems are well suited to demand response as they have the ability to respond, the volume to provide a tangible service to the power system, and the financial incentive to participate in the power market [15].

The demand response capabilities and characteristics of supermarket refrigeration systems are explored through the combined use of time-series modelling and simulation. A statistical model of a supermarket refrigeration system is identified from data procured from the Danfoss refrigeration test centre in Denmark. This model links the power consumption of the refrigeration compressors to representative temperatures of medium and low temperature display units on the system. Full details of this model are provided in [15].

The demand response behaviour of individual flexible loads is simulated in a model predictive control framework, where the refrigeration system tracks a temperature or power reference. The interaction with the aggregator occurs intermittently, where a power reference is requested by the aggregator for a limited period of time. Outside of this period the refrigeration system operates according to its own objectives. Here is assumed the internal objective is to track the average of the temperature limits in the medium temperature display units. The control objective is given as

$$\min \sum_{t=1}^{T} a_t (P_{t}^{ref} - P_t)^2 + (1 - a_t)(T_{t}^{ref} - T_t)^2$$

where the control variables are temperature, $T_t$, and power, $P_t$. A binary indicator, $a_t$, governs the effective control objective at time $t$. When $a_t$ is 1, the aggregator specifies a power reference, $P_{t}^{ref}$ for the refrigeration system to track. When $a_t$ is zero, the supermarket tracks a temperature reference, $T_{t}^{ref}$. The control is subject to upper and lower bounds on temperature in both the medium and low temperature display units. Power consumption is limited by the capacity of the compressors on the system. As the flexibility of this system is defined by the least flexible representative temperature, that of the medium temperature display unit, there is no further reference to the low temperature unit in this work.

The behaviour of the refrigeration system over a period of both supermarket and aggregator control is illustrated in Fig. 1. During this simulation the aggregator requests a reduction in power consumption up to 5kW is requested. The heavy dashed lines indicate the temperature/power references to be tracked, they are active when non-zero.
Under the described control framework, the power consumption of the refrigeration system can be considered to consist of a baseline power consumption and a deviation from this baseline.

In the current model, the baseline power consumption is constant. This is because the system used for model identification is not a fully operational supermarket, and therefore does not include the complexities of customer interaction or widely varying external temperatures. On an operational system the baseline power consumption varies according to a number of factors. This baseline power consumption can be modelled and forecast [21], and purchase of the necessary energy to meet this demand can take place on the day-ahead market. Any deviation from the baseline can be employed to provide regulating power. In order to achieve this it is imperative that the saturation and rebound characteristics are fully described in a manner that can be easily communicated to a power system operator.

III. CHARACTERISING DEMAND RESPONSE

A. Saturation Curve Concept

The time to saturation defines the maximum duration for which any deviation from the baseline can be reliably maintained. This can be found by simulating the response of the system to a range of power adjustments and finding the duration for which the requested power reference can be maintained. The results of these simulations are presented in Fig. 2, which plots the time to saturation against the power adjustment, and shows the closest fit to these points.

The rebound phenomenon can also be described using this curve. If the system is allowed to rebound in an uncontrolled manner, it will tend to do so at either its maximum or minimum power consumption levels, and the duration of this rebound is found at the outer points of the curves. If the aggregator includes a power reference for the rebound, the necessary duration can be found from the corresponding point on the saturation curve. In order to avoid any unexpected saturation or rebound, any service offer from the aggregator to the power system operator must consist of power levels and durations for both response and rebound as defined by the saturation curve. The offer thus has the form of an asymmetric block.

Fig. 3 illustrates the behaviour of the refrigeration system under a request for a response-rebound block consisting of a reduction in consumption by 8.75kW for 145 minutes (response) and an increase in consumption by 12.25kW for 75 minutes (rebound), the block definition shown in Fig. 2. The adjustments occur from a baseline power consumption of 13.75kW. It can be observed from Fig. 3 that the temperature reaches its upper bound, which indicates saturation, and the subsequent rebound is fully controlled. This is achieved without feedback from the supermarket to the aggregator; the aggregator decides on the composition of the entire block before issuing the power references to the supermarket.

The use of the saturation curve to achieve this response illustrates the ability of an aggregator to obtain effective demand response from a single refrigeration unit without the need for detailed modelling, monitoring or communications infrastructures. A similar representation can be found for a population of supermarkets by adding individual saturation curves to form an aggregate curve. The saturation curve of a homogeneous population of supermarkets will have the same form as the saturation curve of an individual supermarket, with a scaled power axis. For example, a population of 1000 supermarkets described by Fig. 2 will have the same saturation curve but the power adjustment will be scaled in MW rather than kW.

B. Saturation Curve Extension

The saturation curve presented in Fig. 2 represents the limits of the demand response capabilities of the refrigeration system. Naturally, the system is also capable of maintaining a power adjustment for a duration less...
than the saturation time, however the necessary rebound following such a response must be defined.

Temperature behaviour within refrigeration units exhibits an exponential relationship with time, for a given power consumption level [16]. However, for small values of $t$ the temperature trajectory is approximately linear. Within the refrigeration systems considered here, the temperature range is relatively small, and the duration for which a given power adjustment can be maintained is limited by saturation. Consequently for values of $\Delta P$ above a given threshold, the duration for which $\Delta P$ is maintained is short and the temperature behaviour in the refrigeration system can be considered linear. This facilitates the identification of partial saturation curves and the definition of the corresponding rebound, if power deviations are only considered outside of a deadband region. This has been verified through simulation for the model considered in this work, where the deadband range is $\{-4, 4\}$ kW.

Consider the extension of the saturation curve concept to incorporate the case where the response is maintained for $X\%$ of the saturation time. An $X\%$ saturation curve can be found for all power adjustments within the linear region by multiplying the original saturation curve by $X$. This facilitates the identification of the appropriate rebound following an $X\%$ response. Fig. 4 illustrates the case where $X = \{25, 50, 75, 100\}$. The advantage of using $X\%$ saturation curves is that the refrigeration units are not stressed to their temperature limits, but instead occupy a limited region around the baseline temperature.

IV. SCHEDULING DEMAND RESPONSE FOR PROVISION OF REGULATING POWER

A. Problem Context and Assumptions

Demand response units are scheduled alongside conventional generating units for the provision of regulating power. The system dispatch is subject to two key simplifying assumptions. Firstly, that the system operator has perfect foresight of the required regulating power within the considered horizon. Secondly, all conventional generating units can provide up- and down-regulation, and the maximum amounts of each are fixed for the duration of the optimisation. It is assumed that their existing dispatch (e.g. from the day-ahead market) allows for this.

B. Problem Formulation

The optimal dispatch of conventional and demand response units is found by employing the mixed integer linear programming optimisation given as

$$\min x^T c$$

subject to:

$$h(x) = 0$$

$$g(x) \geq 0$$

where $x = \{P_{d,c}, P_{d,dr}, v_{d,c,t}, S_{d,c,t}, SD_{d,c,t}\}$, the conventional generator power output of each generating unit, $i$; the demand response power output for each block, $c$; and unit, $d$; the online status of the demand response block, $c$ at unit $d$; and its start-up and shut-down indicators respectively. The objective function, (2a), minimises the cost to the system operator of sourcing regulating power subject to the sets of equality and inequality constraints governing the generating and demand response units on the power system. The generating unit constraints are those typically employed in economic dispatch and can be found in a number of references, including [22].

The constraints governing the behaviour of the demand response units are provided in equations (3) - (6). The initialisation and conclusion of a demand response block are indicated by a change in the online status of a given block, $v_{d,c,t}$, as detailed in (3).

When a demand response block is requested by the system operator, the demand response unit must follow the profile of the block, as defined in (4). This profile is comprised of the response, $P^{DR, resp}_{d,c}$, and rebound, $P^{DR, reb}_{d,c}$, for the corresponding response and rebound durations, $T^{resp}_{d,c}$ and $T^{reb}_{d,c}$. There may also be a recovery period following the completion of a demand response block, $T^{rec}_{d,c}$. Each demand response unit offers a number of demand response blocks, however simultaneous activation of blocks from a single demand response unit is not allowed. This is imposed in (5). Finally, any activated block must be fully realized within the dispatch horizon. This constraint is enforced in (6) which ensures that demand response blocks cannot commence in the final periods of the dispatch window, where this restricted region is defined by the response and rebound durations of each block. These constraints are summarised as

$$v_{d,c,t} - v_{d,c,t-1} = SU_{d,c,t} - SD_{d,c,t},$$

Fig. 4. Partial saturation curves with the dead-band indicated by the shaded grey section. To ensure accuracy, partial saturation curves should not be considered for power adjustments within the shaded region.
the rebound portion of the block. Equation (8c) imposes a
minimum recovery period, $T_{d,c}^{rec}$, between the activation
of blocks from unit $c$. This constraint ensures that no
block is active (i.e. $v_{d,c,t} = 0$) for the recovery period
following the response and rebound, given that a block
has been activated at time $t$. This implementation is
given as

$$
0 \leq 0 - (1 - SU_{d,c,t})M, \\
\text{if } \alpha_{d,c} = 1 \\
0 \geq 0 - (1 - SU_{d,c,t})M, \\
\text{if } \alpha_{d,c} = 0
$$

V. CASE STUDY DEFINITION

Case studies are employed in this work to demonstrate
the value of demand response to the system operator
when its abilities are described using the limited form of
a response-rebound block. Demand response is consid-
ered for the provision of regulating power for the Belgian
regulating power market. Three cases are considered:

1) Dispatch of the system without demand response.

2) Dispatch of the system considering a limited set of
demand response block offers.

3) Dispatch of the system considering a fully observ-
able and controllable demand response resource.

Historical regulating power data from the Belgian
system operator, ELIA, is employed in all case studies.
On this power system, regulating power is recorded at
a 15 minute resolution. The data is interpolated to 5
minute resolution using cubic splines to match the time
resolution of the demand response models. Each case
study considers a dispatch window of 4 hours, using
data from 2012.

The demand response resource consists of two de-
mand response units, which each consist of a population
of flexible loads. The flexibility of each unit is described
using six response-rebound block offers, as detailed in
Table I. While the physical capabilities of the units are
the same, different blocks are offered for dispatch. This
reflects the expectation that in a real-world implementa-
tion supermarkets would be clustered together to offer
different services to the system operator. The blocks are
selected from the 50% saturation curve shown in Fig. 4.

For comparative purposes, the demand response re-
source is also implemented in its fully observable and
controllable form, as a time series model. The time series
model is that from which the saturation characteristic
and block offers are determined. Dispatch of the fully
modelled units is subject to the restriction that they must reach the mean of their temperature bounds at the end of the dispatch horizon. This ensures an approximate energy balance and a fair comparison between the full and limited representations of the demand response resource.

Each demand unit has a maximum up-regulating capacity of 13 MW and down-regulating capacity of 17 MW. The capacity of the demand response units is scaled to be comparable to the capacity of the conventional units considered in the case study. This scaling can be interpreted to represent a homogeneous population of 1000 individual supermarkets. The cost of acquiring up- and down-regulation from the demand response units is set at 2EUR/MWh. This is less than the cost of sourcing regulating power from any of the conventional units, ensuring that the demand response resource will be first in the merit order.

Four conventional generating units are considered in the case studies. Table II contains the technical specifications of each unit and the costs of acquiring up- and down-regulation from each. These costs are based on the production cost of each unit, where the up- and down-regulating costs are the production costs multiplied by a factor greater than and less that one respectively. The scaling factors are found through an analysis of the difference between the day-ahead price on the Nordic power market and the up- and down-regulating prices [13].

Six regulating power profiles are employed in the case studies to evaluate the demand response resource, as shown in Fig. 5. Case A comprises 3 slowly varying profiles, while Case B comprises 3 fluctuating profiles. It is expected that the demand response blocks will have greater value in situations where the regulating power requirement fluctuates due to their asymmetric shape and the large effective ramp rate between the response and rebound portions of the block. The two sets of regulating power profiles are considered for comparison. It is the experience of the authors from sourcing these profiles that Case B better represents typical conditions than Case A.

VI. RESULTS AND DISCUSSION

The case study results are presented in Tables III and IV. The theoretical value of demand response is defined as the amount by which the cost of meeting regulating requirements is reduced when demand response is represented using a fully observable and controllable model. This is compared to the practical value that this resource can provide to the system when represented by a limited set of blocks. It is evident that demand response is capable of providing substantial value to the system, and as expected there is a significant difference between the theoretical and practical resources.

This difference is due to two key factors. Firstly, the block definition imposes the need for a response and rebound that directly follow one another. This differs from the operation using the full model of the refrigeration system, where the only restriction is that an approximate energy balance is maintained (as imposed with a final temperature constraint). This allows the response and rebound to be separated. Consequently, the unit has greater flexibility to follow the regulating power profile rather than requiring a rebound which may be in the opposite direction to the required regulating power. Secondly, the block definitions are formulated using the 50% saturation curve, which has an effective temperature range of approximately 50% of the full range using the absolute limits imposed by the supermarket operator. In comparison, the fully modelled demand resource is free to employ the full temperature range, resulting in greater overall flexibility. Imposing tighter temperature limitations on the fully modelled resource allows the
Comparison of the value of both the block definitions and the full model when they are operating with the same physical flexibility. This comparison is included in the last row of Table IV, where it is observed that the disparity between the two forms of demand response is significantly reduced. This indicates that a very limited representation of the demand response capabilities of this thermal system has comparable value to a fully described system. The cost of establishing, controlling, monitoring and operating a fully modelled system is very high, and this result indicates that such a cost may not be justified by the additional value it brings to system operation.

Comparison of Tables III and IV reveals that there is a greater disparity between the theoretical and practical values when the regulating power profiles vary slowly, as in Case A. Analysis of the behaviour of the fully modelled units in Cases A1-A3 reveals that they tend to provide both response and rebound in the prevailing direction of the regulating power profile. In contrast, the blocks require a rebound immediately following the response. This must be compensated for by conventional units if the rebound is in the opposite direction to the required regulating power. Consequently, scheduling blocks for slowly varying regulating profiles is either very costly, or the blocks are not scheduled at all. This is confirmed in Table III, where a larger difference between the theoretical and practical values is observed in case A1 than in any of the B cases, and in case A3 where the value of demand response described using blocks is negligible. In case A2, the demand response is incapable of bringing any significant value to the system, regardless of the resource description used. This is because the regulating power requirement is very high so the percentage contribution from demand response is lower than in the other cases.

Table IV presents the cost reductions for regulating power procurement when the demand response resource is represented with a varying number of blocks. For cases with less than 6 blocks, the blocks are taken in order from Table I. It can be concluded from Table IV that the value of the demand response resource when described using block offers approaches the value of the fully modelled resource as the numbers of block offers increases. In fact, if the flexibility of the demand response resource were described using an infinite number of block offers, it would be equivalent to the flexibility described by the fully modelled system.

It is shown in Table IV that in some cases, increasing the number of blocks has no impact on costs. This is because the additional block offer is not selected for dispatch, and can be understood to be unsuitable for the considered regulating power profile. The results of this analysis reveal that cost savings greater than 20% can be achieved with only four block offers. This demonstrates that even a very limited representation of the flexible demand resource facilitates significant cost savings.

Fig. 6 illustrates the aggregate dispatch of the generating and demand response units for case B1. The most beneficial behaviour in terms of system dispatch cost would be for the demand response blocks to reduce the power provided by generating units. This behaviour is observed for the majority of the dispatch horizon, however there are brief periods where the generating units are required to compensate for the rebound of the demand response units. This can be observed during the interval between minutes 145 and 160. During this interval one of the demand response units is rebounding in the opposite direction to the required regulating power and the generating units must provide additional down-regulation. From minute 165 the second demand response unit begins providing down-regulation which partially compensates for the rebound of the first unit and reduces the over-provision from the generating units. Despite this need for compensation, the demand response blocks offer significant value to the system when optimised for cost minimisation. In the case of a volume-based optimisation, this form of demand response may not be attractive.

Fig. 7 illustrates the dispatch of the demand response blocks and the fully modelled resource for case B1. It is evident that the demand response blocks attempt to

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>COST REDUCTION WITH DEMAND RESPONSE - CASE A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>6 DR Block Offers</td>
<td>10.53%</td>
</tr>
<tr>
<td>Fully Modelled Demand</td>
<td>36.63%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>COST REDUCTION WITH DEMAND RESPONSE - CASE B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B1</td>
</tr>
<tr>
<td>2 DR Block Offers</td>
<td>9.54%</td>
</tr>
<tr>
<td>3 DR Block Offers</td>
<td>17.10%</td>
</tr>
<tr>
<td>4 DR Block Offers</td>
<td>20.81%</td>
</tr>
<tr>
<td>5 DR Block Offers</td>
<td>21.23%</td>
</tr>
<tr>
<td>6 DR Block Offers</td>
<td>21.43%</td>
</tr>
<tr>
<td>Fully Modelled Demand</td>
<td>36.78%</td>
</tr>
<tr>
<td>Limited Temperature Range</td>
<td>24.45%</td>
</tr>
</tbody>
</table>
dispatch demand response block offers, which require binary variables that are computationally burdensome for large scale implementation.

REFERENCES