Integration of Flexible Consumers in the Ancillary Service Markets
Biegel, Benjamin; Westenholz, Mikkel; Hansen, Lars Henrik; Stoustrup, Jakob; Andersen, Palle; Harbo, Silas

Published in:
Energy

DOI (link to publication from Publisher):
10.1016/j.energy.2014.01.073

Publication date:
2014

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: december 17, 2018
Integration of flexible consumers in the ancillary service markets

Benjamin Biegel\textsuperscript{a,}\textsuperscript{*}, Mikkel Westenholzb, Lars Henrik Hansenc, Jakob Stoustrup\textsuperscript{a}, Palle Andersen\textsuperscript{a}, Silas Harbo\textsuperscript{d}

\textsuperscript{a}Department of Electronic Systems, Automation and Control, Aalborg University, Denmark
\textsuperscript{b}Better Place, Copenhagen, Denmark
\textsuperscript{c}DONG Energy AS, Copenhagen, Denmark
\textsuperscript{d}Danish Energy Association, Copenhagen, Denmark

\section{Introduction}

The renewable energy sector is the fastest growing power generation sector and is expected to keep growing over the coming years \cite{1,2}; the global share of non-hydro renewables has grown from 2\% in 2006 to 4\% in 2011 and is predicted to reach 8\% in 2018 \cite{2}. Many actions have been taken all over the world to increase the penetration of renewables: in the US, almost all states have renewable portfolio standards or goals that ensure a certain percentage of renewables \cite{3}; similarly, the commission of the European Community has set a target of 20\% renewables by 2020 \cite{4}.

A number of challenges arise as the penetration of renewables increases. Many renewable sources are characterized by highly fluctuating power generation and can suddenly increase or decrease production depending on weather conditions. A recent example of this phenomenon took place Denmark on October 28, 2013 where a large number of wind turbines were shut down because of a storm. This caused a decrease from a level where more than 100\% of the Danish electricity consumption was covered by wind to a level less than 45\% in just 2 h,\textsuperscript{1} see Fig. 1. Such rapid production changes can imply severe consequences for grid stability due to the difficulty of accurately predicting the timing of the events \cite{6}.

Further, as more renewables are installed, the conventional generators are phased out: in Denmark, the increase of renewables during the last years has caused a petition for shutting down 8 central power plants \cite{7}. This, however, causes another major challenge because the central power plants currently are the providers of system stabilizing ancillary services. As the conventional power plants are replaced with renewables, the ability to provide ancillary services in the classical sense is lost as the renewables usually do not possess the ability to provide such system stabilizing reserves: First of all, keeping renewables in reserve will entail that free energy is wasted making this a very expensive solution. Second, the highly fluctuating nature of the renewables caused by

\begin{itemize}
  \item \textsuperscript{1} Data taken from the website of the Danish transmission system operator: Ref. \cite{5}.
\end{itemize}
weather conditions can make it difficult to deliver a well-defined power response.

It is therefore evident that alternative sources of ancillary services must be established as renewables replace conventional generation. One approach to obtain ancillary services is to purchase reserves in neighboring countries; however, this requires that transmission line capacity is reserved for the reserve markets which will limit the capacity in the day-ahead spot markets and thereby possibly cause higher electricity prices [7]. Further, the ENTSO-E (European network of transmission system operators for electricity) grid code sets limits on the amount of reserves it is allowed to exchange internationally [8].

An alternative approach to obtain alternative ancillary services is the smart grid concept, where local generation and demand-side devices with flexible power consumption take part in the balancing effort [9,10]. The basic idea is to let an aggregator control a portfolio of flexible devices such as thermal devices, batteries, pumping systems etc. Hereby, the aggregator can utilize the accumulated flexibility in the unbundled electricity markets for primary, secondary, and tertiary reserves, on equal terms with conventional generators [11,12].

In this work, we identify the difficulties of including flexible consumption devices in the existing ancillary service markets and propose a method for better integration of this type of devices.

2. Scope and structure of the article

The increase of renewables and shutdown of central power plants call for alternative sources of primary, secondary, and tertiary reserves. This work proposes a method for making better conditions for flexible consumption devices to deliver these services. The method is valid for both the primary and secondary reserve, but not for the tertiary reserve, as will be come evident later. For the following reasons, we still believe the method is most relevant.

The first reason is that flexible consumption devices and storage systems are well suited for fast reserves but less suited for slower reserves where large amounts of energy must be delivered. Many consumption devices are able to deliver a response fast enough even for primary reserve [13,14]; however, they are not able to provide actual energy deliveries as they only have a limited energy capacity. A battery system will for example only be able to deliver/consume a limited amount of energy before reaching the energy limitations; similarly, a consumption devices with a given thermal mass will only be able to shift a limited amount of energy before reaching the thermal comfort limits [12].

The second reason is that although the amounts of required tertiary reserves is significantly higher than the required amount of primary and secondary reserves, the expenditure on primary and secondary reserve exceeds that of tertiary reserve by far. This is illustrated in Fig. 2 where 2011 and 2012 data for Western Denmark is analyzed. The figure shows that the amount of tertiary reserve in 2011 and 2012 indeed is the highest of the tree comprising more than 50% and 55%, respectively, of the combined primary, secondary and tertiary reserve those years. However, as illustrated in the same figure, the expenditure for the tertiary reserve in these two years accounted for below 12% and 11%, respectively. The reason is the fast delivery requirements for primary and secondary reserves making it more difficult, and thus more costly, to provide these reserves.

Based on the observation that flexible consumers are well suited for fast reserves and because the value of these services is far greater than of tertiary reserve, it is chosen to limit the scope exclusively to primary and secondary reserves.

A portfolio of flexible consumption devices generally has two significant differences from conventional power generators when providing ancillary services. The first is that the portfolio will have a limited energy capacity whereas the conventional generator simply will be able to use more or less fuel. A heating system will for example have flexibility due to its thermal capacity; however, only a limited amount of energy can be stored depending on the temperature bounds that must be satisfied. Similarly, a factory may be able to expedite or postpone a batch production, but will in the long run have the same average consumption. This significantly limits the possibilities for flexible consumption devices to provide ancillary services. The second difference is that a portfolio of flexible

---

2 Data for primary and tertiary reserve taken from Ref. [5] while data for secondary reserve is from Ref. [13,16]. Only the reservation prices are included, not the activation prices which only apply for secondary and tertiary reserves.
devices often not will have a well-defined baseline, i.e. the aggregator will not exactly know the electricity consumption of the portfolio many hours in advance as it depends on external parameters such as weather conditions or human behavior, which can be difficult to predict accurately. Without a well-defined baseline it is difficult to assess what services the portfolio actually has delivered; consequently, the lack of a baseline makes it difficult for flexible consumers to participate in the ancillary service markets under the current regulations. These two issues therefore constitute a barrier for the roll out of the smart grid concept in the liberalized electricity markets.

In this work, we propose a method that resolves the issues of energy limitations and lack of accurate baselines without altering the existing ancillary service markets. In short, the method allows an aggregator via ICT (information and communication technology) to continuously adjust its operational schedule which is the baseline communicated to the TSO (transmission system operator). This enables the aggregator to avoid violating the energy limitations of the consumption devices. The operational schedule adjustments must, however, be done under certain limitations ensuring that the TSO has sufficient time to activate slower reserves correspondingly.

The proposal is exactly in line with the general smart grid vision where a stable, reliable, and sustainable electricity system is ensured via ICT solutions [11,17,18].

The paper is organized as follows. First in Sec.3 we describe the overall system architecture. Following in Sec. 4 and Sec. 5 we present overall models of flexible consumption devices and of the ancillary service markets, respectively. In Sec. 6 we discuss the issues of delivering ancillary service via flexible consumers and following in Sec. 7 we present our proposal of resolving these issues. The proposed method is illustrated with numerical examples in Sec. 8 and finally in Sec. 9 we conclude the work.

3. Architecture

For many consumption devices, the flexibility is too small to make isolated bids into the electricity markets; for example, the threshold for primary frequency control reserves is 300 kW in Western Denmark [19] while the capacity of a domestic flexible consumption device is in the magnitude of a few kW at most. Only very large consumers such as large pumping facilities, heating elements for combined heat and power plants, etc. will be able to reach the minimum threshold. For this reason, aggregation is required in order to achieve sufficient quantities of active power for bidding.

The basic idea is to let an aggregator enter into contract with the owners of the flexible devices. The contract specifies under what conditions the aggregator is allowed to utilize the flexibility [20]. On this basis, the aggregator uses a technical unit often referred to as a VPP (virtual power plant) to manage the devices [12]. The VPP can monitor and control the flexible devices and is thereby able to mobilize the accumulated response of a portfolio of flexible consumption devices, see Refs. [21–24] for a few examples of VPP strategies. This allows an aggregator to enter the ancillary service markets based on the flexible devices. This architecture is illustrated in Fig. 3.

4. Flexible consumption devices and storage devices

In this section, we present a model that describes a portfolio of flexible consumption devices managed by an aggregator. The model is very simple but captures characteristics in focus in this work: power and energy limitations and inaccurate knowledge of the consumption baseline.

4.1. Nomenclature

Table 1 gives an overview of the parameters used in the following modeling section. Later, each parameter is described in more detail; further, some of the parameters are illustrated in Fig. 4.

4.2. Model

A flexible consumption device portfolio model can be described as follows. Let \( E(t) \) denote an energy level and define its derivative as

\[
\dot{E}(t) = P_{\text{cons}}(t) - P_{\text{base}}(t)
\]

where \( P_{\text{cons}}(t) \) is the portfolio electricity consumption \( P_{\text{base}}(t) \), is the baseline consumption of the portfolio, i.e., how much the portfolio of devices would consume if not activated for ancillary services, and \( E(t) \) is the energy stored in the flexible consumption devices. In other words: by deviating from the nominal portfolio baseline consumption \( P_{\text{base}}(t) \), energy is stored or released from the portfolio. Notice that the baseline consumption always will be non-negative \( P_{\text{base}}(t) \geq 0 \) as the portfolio does not include power generators.

The model (1) can also be utilized for a battery storage. In this case the baseline consumption will simply be zero \( P_{\text{base}}(t) = 0 \) whereby \( E_{\text{batt}}(t) = P_{\text{cons}}(t) \), given the battery is not used for other purposes and does not have any drain/loss. Now, as the battery charges we will have \( P_{\text{cons}}(t) > 0 \) and the battery level \( E_{\text{batt}}(t) \) will increase and vice versa for discharge.

The consumption of the portfolio is limited in power and energy, which can be represented as

\[
p_{\text{min}} \leq P_{\text{cons}}(t) \leq P_{\text{max}}, \quad E_{\text{min}} \leq E(t) \leq E_{\text{max}}
\]

where \( p_{\text{min}}, p_{\text{max}} \) represent the limits of the portfolio’s accumulated consumption. For a portfolio of consumption devices \( p_{\text{min}} \), could be 0 if it is allowed to turn all devices OFF; similarly \( p_{\text{max}} \), could be the total consumption with all devices ON, provided this is allowed. For a battery system \( p_{\text{min}}, p_{\text{max}} \), will correspond to the maximum rate of charge and discharge. The parameters \( E_{\text{min}}, E_{\text{max}} \) are the minimum and maximum amount of stored energy and can for example represent an allowable temperature band for thermal devices;

\[\text{Notice that storing electricity for consumption devices refer to the device’s ability to shift consumption in time within certain limits.}\]
similarly, it can represent the limits of a battery. Notice that for consumption devices we will have $P_{\text{min}} \geq 0$ as the devices cannot generate electricity.

Note that modeling a portfolio of many individual devices with a single lumped model as the one presented above in many cases is a vast simplification of reality [25]. Further note that the model does not account for state dependent losses, i.e. it can for example not capture that the energy loss of a thermal device will increase with increasing temperature difference to the ambient. Consequently, the presented model is a rough estimation of reality. However, the focus of this work is not modeling but rather the proposal for a market change that can increase the market uptake of flexible consumers. As the presented model is able to capture the main characteristics of flexible consumers, namely energy limitations and inaccurate baseline predictions, the model is found suitable for this work.

Based on (1) and (2) we define the power capacity $P_{\text{cap}}$ of the portfolio within a specific delivery time $T$ as

$$P_{\text{cap}} = \min\left( p_{\text{max}} - \max_{t \in T}(P_{\text{base}}(t)), \min_{t \in T}(P_{\text{base}}(t)) - P_{\text{min}} \right).$$

(3)

Hereby, the power capacity describes the maximum possible deviation in either direction away from the power baseline within the horizon $T = \{ t \in \mathbb{R} | 0 \leq t \leq T \}$. The basis of this definition is the underlying assumption that the portfolio as default will consume the baseline consumption and deviate from this baseline upon ancillary service activation. In this case $P_{\text{cap}}$ is the highest symmetric power bid we can make. By symmetric, we mean that when a reserve capacity of size $P_{\text{cap}}$ is sold, the provider should be able to deliver power within the symmetric interval $[-P_{\text{cap}}, P_{\text{cap}}]$. This illustrates that flexibility of a portfolio is highest when the baseline consumption is constant and given by $P_{\text{opt}}^{\text{base}}(t) = (P_{\text{max}} - P_{\text{min}}) / 2$ whereby $P_{\text{opt}}^{\text{base}} = (P_{\text{max}} - P_{\text{min}}) / 2$. For the energy part, we define the capacity $E_{\text{cap}}$ as the size of the energy storage:

$$E_{\text{cap}} = P_{\text{max}} - E_{\text{min}}.$$  

(4)

The baseline consumption $P_{\text{base}}^{\text{opt}}(t)$ can be predicted with a given accuracy for a given horizon. Let $P_{\text{base}}^{\text{opt}}(t)$ denote the prediction of $P_{\text{base}}(t)$ and let the accuracy of the prediction be described as

$$|P_{\text{base}}^{\text{opt}}(t) - P_{\text{base}}(t)| \leq P_{\text{acc}}, \ \forall t \in T$$

(5)

where $P_{\text{acc}}$ represents the accuracy. The parameter $P_{\text{acc}}$ can for example describe the ability to predict the outdoor temperature which is relevant when dealing with a portfolio of heating or cooling devices, or it can describe disturbances such as human behavior which is relevant for heating systems of households.

It is necessary for the aggregator to report an operational schedule $P_{\text{op}}^{\text{t}}(t)$ to the TSO describing the scheduled portfolio consumption. The operational schedule must be submitted day-ahead and describes the consumption of the portfolio the following day with a given resolution. As an example, the deadline for the operational schedule is 17.00 in the Danish market and the resolution is 5 min [26]. The aggregator can for example choose to assign the predicted baseline consumption as the operational schedule $P_{\text{op}}^{\text{cons}}(t) : = P_{\text{base}}^{\text{opt}}(t)$ as this is the best possible prediction of the actual baseline consumption $P_{\text{base}}(t)$.

By definition, ancillary services are delivered by letting consumption deviate from the operational schedule. If we let $P_{\text{del}}^{\text{t}}(t)$ denote the delivered ancillary service, we have

$$P_{\text{del}}^{\text{t}}(t) = P_{\text{op}}^{\text{cons}}(t) - P_{\text{cons}}^{\text{t}}(t).$$

(6)

where $P_{\text{del}}^{\text{t}}(t)$ is in production terms, i.e. $P_{\text{del}}^{\text{t}}(t) > 0$, corresponds to increased production or reduced consumption while we use consumption terms for $P_{\text{cons}}^{\text{t}}(t)$, $P_{\text{del}}^{\text{t}}(t)$, $P_{\text{base}}^{\text{t}}(t)$, i.e. $P_{\text{cons}}^{\text{t}}(t) > 0$, corresponds to consuming power. The complete setup is illustrated in Fig. 4.

In this work we propose that an aggregator is allowed to adjust its operational schedule as long it is done sufficiently slowly, such that the TSO is able to activate slower reserves accordingly. This allows the aggregator to keep the energy level of its portfolio close to a certain desired level, e.g. for the energy midpoint $E_{\text{min}} - E_{\text{cap}} / 2$, and hereby avoid violating the energy limits.

### 4.3. Examples

Let us consider a few concrete examples of flexible consumption devices that are considered potential providers of ancillary services in the smart grid literature.

The first example is a household heated with a heat pump which can be seen as a flexible consumption device due to the thermal mass of the house [27–29]. The energy/power parameters will vary much from house to house. To give an example, a set of parameters for a smaller house where we are allowed to vary the temperature a few degrees around the temperature set-point is presented in Table 2 inspired by the papers cited above.

The second device is a supermarket refrigeration system where energy can be stored in the refrigerated foodstuff [30–32]. A set of parameters for a smaller supermarket system where we are allowed to lower the foodstuff temperature a few degrees is presented in Table 2 inspired by Ref. [33].

---

### Table 1: Description and units of the parameters used throughout the work.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(t)$</td>
<td>Energy level in portfolio</td>
<td>[J]</td>
</tr>
<tr>
<td>$E_{\text{min}}, E_{\text{max}}$</td>
<td>Portfolio min/max energy levels</td>
<td>[J]</td>
</tr>
<tr>
<td>$P_{\text{min}}, P_{\text{max}}(t)$</td>
<td>Portfolio power consumption</td>
<td>[W]</td>
</tr>
<tr>
<td>$P_{\text{base}}(t)$</td>
<td>Portfolio baseline consumption</td>
<td>[W]</td>
</tr>
<tr>
<td>$P_{\text{cap}}$</td>
<td>Portfolio power capacity (largest possible symmetric power bid)</td>
<td>[W]</td>
</tr>
<tr>
<td>$E_{\text{cap}}$</td>
<td>Portfolio energy capacity (maximum amount of energy that can be stored)</td>
<td>[J]</td>
</tr>
<tr>
<td>$P_{\text{base}}^{\text{opt}}(t)$</td>
<td>Prediction of the baseline consumption $P_{\text{base}}^{\text{opt}}(t)$</td>
<td>[W]</td>
</tr>
<tr>
<td>$P_{\text{acc}}$</td>
<td>Accuracy of baseline prediction within horizon</td>
<td>[W]</td>
</tr>
<tr>
<td>$P_{\text{del}}^{\text{t}}(t)$</td>
<td>Amount of symmetric reserve</td>
<td>[W]</td>
</tr>
<tr>
<td>$P_{\text{op}}^{\text{t}}(t)$</td>
<td>Operational schedule reported to the TSO</td>
<td>[W]</td>
</tr>
</tbody>
</table>

---

![Fig. 4. Illustration of the simple model of a portfolio of flexible consumers and how it is able to make a power delivery $P_{\text{del}}^{\text{t}}(t)$ by deviating from the operational schedule $P_{\text{op}}^{\text{t}}(t)$.](image-url)
Finally we also consider an EV (electrical vehicle) battery. Typical values for an EV battery are presented in Table 2 [34,35]. We assume a fast DC (direct current) charging station and that the battery is not in use, which would be the case for example if the battery is located at a charging station.

These examples are presented to illustrate the types of devices that go under the category flexible consumption devices in this work and to give an idea of the energy and power capacities of such devices.

We notice, as previously mentioned, that all these devices are too small for individual participation in the ancillary service markets where the threshold is 300 kW or more; consequently, aggregation is a requirement.

5. Ancillary service markets

We limit our focus to the active power ancillary services although other ancillary services exist. The active power services are denoted primary, secondary, and tertiary reserve as previously mentioned. In ENTSO-E’s network code on load-frequency control and reserves, the terminology used for these services are frequency containment reserve, frequency restoration reserve, and replacement reserves [8,36]. These terms describe the functionality of the reserves in case the system frequency deviates from the nominal value: namely that the fast primary reserve ensures that the frequency is contained, the secondary reserve restores the frequency, while finally, the tertiary reserve replaces the secondary reserve.

We assume these services are distinguished by how fast they are with primary as the fastest and tertiary as the slowest reserve. In this work we describe a method that allows an aggregator providing fast reserves, for example primary reserve, to utilize the slower reserves, for example secondary reserve, to ensure that the energy limitations of the portfolio are not violated.

Throughout the examples, we examine providing primary reserve and utilizing the markets for secondary or tertiary reserve to restore the portfolio energy level; however, the method would also apply to a case where we provide secondary reserve and utilize the market for tertiary reserve to restore the portfolio energy level.

5.1. Generic market description

In the following we construct a simple description of the active power reserves seen from an ancillary service provider’s point of view.

A provider has contracted a capacity given by \( P^{\text{res}} \) for a duration given by \( T_i \); where the subscript \( i \) denotes the market, i.e. \( i = 1 \) is the primary \( i = 2 \), is the secondary, and \( i = 3 \) is the tertiary reserve market. This notation is used throughout this work. For simplicity we only consider symmetric deliveries.

We use the following simple model to describe the ancillary service markets: each market \( i \) is described by two parameters: a ramping time \( t^{\text{ramp}}_i \) and a latency time \( t^{\text{lat}}_i \). These parameters should be understood as follows. If a reserve is fully activated, either via local grid frequency measurements for primary reserve or by activation from a TSO for secondary and tertiary reserve, the provider should start providing the reserve at the latest after the latency time of \( t^{\text{lat}}_i \) seconds; hereafter the full reserve should be ramped up within an additional \( t^{\text{ramp}}_i \) seconds.

Generally, primary control needs faster response than secondary control which needs faster response than tertiary control. This can be described in terms of the ramping and latency parameters:

\[
\begin{align*}
t_1^{\text{ramp}} & \leq t_2^{\text{ramp}} \leq t_3^{\text{ramp}} \\
t_1^{\text{lat}} & \leq t_2^{\text{lat}} \leq t_3^{\text{lat}}.
\end{align*}
\]

The faster reserves are in average more expensive than the slower, as they are more difficult to deliver. This is the reason it is interesting to examine how flexible consumption devices can be managed to deliver the fast expensive reserves by restoring the energy level via the inexpensive slower reserves.

5.2. Example: European grid

We consider a concrete example by examining the control performance specifications of the ENTSO-E. Based on [37,38] as well as the newly published grid code [8], typical parameters for the three ancillary services are

\[
\begin{align*}
t_1^{\text{ramp}} &= 30 \text{ s}, & t_1^{\text{lat}} &= 0 \text{ s} \\
t_2^{\text{ramp}} &= 6 \text{ min}, & t_2^{\text{lat}} &= 30 \text{ s} \\
t_3^{\text{ramp}} &= 10 \text{ min}, & t_3^{\text{lat}} &= 5 \text{ min}.
\end{align*}
\]

The parameters stated in (8) should not be seen as definite values as they can vary from country to country, but they are chosen to mimic the parameters presented in ([38], p. 3).

An illustration of primary, secondary and tertiary reserve can be seen in Fig. 5 with the parameters from (8) and assuming an instance of 1 MW at time 0. Further, it is assumed that the fault is corrected by three providers of each 1 MW reserve, i.e. \( P^{\text{res}}_1 = P^{\text{res}}_2 = P^{\text{res}}_3 = 1 \text{ MW} \). The figure shows that the primary response within 30 s fully provides the 1 MW of power where after the secondary reserve starts ramping up followed by the tertiary reserve after another 4.5 min. The secondary reserve thus restores the primary reserve and is itself eventually restored by the tertiary reserve.

Further we notice that the figure illustrates what was discussed in Sec. 2, namely that the required amount of tertiary reserve is

![Fig. 5](image-url) A 1 MW instance is restored by the primary reserve which is relieved by the secondary reserve which again is relieved by the tertiary reserve.
larger than that of primary and secondary reserve because the tertiary control will replace the primary and secondary control action and provide an actual energy delivery. The figure also illustrates the higher timing requirements to the primary and secondary control action, which is the reason for the higher absolute costs of these reserves although the volumes are smaller, as was illustrated previously in Fig. 2.

Finally we comment on the delivery duration $T_i$ of the reserves which is the duration that the contracted reserve $P_{f}^{\text{res}}$ should be available. The delivery duration vary from market to market, however we use the Danish system as an example [39]

$$T_1 = 1 \text{ week}^4, \quad T_2 = 1 \text{ month}, \quad T_3 = 1 \text{ hour}. \quad (9)$$

This means that the reserves are sold in blocks of 1 week, 1 month, and 1 h, respectively.

Finally, we look at the ancillary service prices in Denmark to illustrate that the faster reserves are more expensive than the slower reserve. Let $\pi_i$ denote the average cost per MW of reserve, then

$$\pi_1 = 30 \text{ \varepsilon/MW}, \quad \pi_2 = 11 \text{ \varepsilon/MW}, \quad \pi_3 = 5 \text{ \varepsilon/MW}. \quad (10)$$

where these prices are taken from Ref. [19] for the secondary reserve and based on prices from the first 6 months of 2013 for the primary and tertiary reserve.

6. Ancillary services by flexible consumers

The limiting factors for conventional generators to provide ancillary services are their power limitations, the startup time, and ramping limitations. Generally, the energy capacity of a conventional generator is a non-issue: the generator will be able to continuously produce both minimum and maximum power simply by using more or less fuel.

For flexible consumers the situation is completely different. Consumption devices will typically hardly have any ramping limitations and have a very low startup (or shutdown) time. The reason is that the consumption devices often rapidly can change the process to consume more or less power or it can simply be turned ON/OFF and thus instantaneously change the power consumption. This makes flexible consumption devices ideal for providing fast reserves such as primary reserve. This further illustrates why it is very interesting to improve the possibility for these devices to participate in the fast ancillary service markets.

As previously described, two main differences from conventional generators make it difficult for flexible consumption devices to provide ancillary services: First, the flexible consumption devices are energy-limited and they will therefore on average have to consume the same energy and consequently not be able to provide actual energy deliveries. Second, the flexible consumption devices generally do not have an exact baseline for the future consumption. In the following we will describe why this becomes a limiting factor for the flexible consumption devices as providers of ancillary services in the current markets.

6.1. Energy limitations

It is easy to illustrate how the energy limitations can limit the power delivery $P_{f}^{\text{res}}$ we are able to offer as an aggregator. An aggregator providing ancillary services in market $i$ should in principle be able to deliver the reserve within the power limitations $\pm P_{f}^{\text{res}}$ continuously throughout the delivery period $T_i$. For the primary reserve market with a duration of one week, this means that the worst case energy deliver in principle is $\pm 168 \text{ hours} \cdot P_{f}^{\text{res}}$.

As an example, a portfolio of 100 EV batteries with an energy capacity of $E^{\text{cap}} = 2.4 \text{ MWh}$ can at most bid a symmetric power reserve of $P_{1}^{\text{res}} = 2.4 / (2 \cdot 168) = 0.007 \text{ MW}$ which is very restrictive compared to the power capacity of $P^{\text{cap}} = 5.0 \text{ MW}$. It can be argued that in practice, an extreme energy delivery of $\pm 168 \text{ hours} \cdot P_{2}^{\text{res}}$ will not occur. However, by examining historical grid frequency measurements, weeks can be found where an energy delivery in the magnitude of $\pm 10 \text{ hours} \cdot P_{2}^{\text{res}}$ is required. This yields $P_{2}^{\text{res}} = 2.4 / (2 \cdot 10) = 0.12 \text{ MW}$ which is still very low compared to the 5.0 MW power capacity available.

Notice that the restriction depends on how the energy and power capacity relates: the problem increases for a portfolio with a relatively high power capacity compared to energy capacity. For a portfolio of heat pumps or supermarket refrigeration systems, the issue is smaller than the example presented above, however it will be worse for other types of devices with even smaller energy capacities, for example thermal devices with very tight allowable temperature bands. Finally, the situation will be even worse for the secondary reserve where the duration is longer, namely $T_2 = 1$ month.

6.2. Uncertain baseline

It is also easy to illustrate how the uncertain baseline can be a limiting factor for how large a power delivery $P_{f}^{\text{res}}$ we are able to offer based on a portfolio of flexible consumption devices. The provided reserve is defined as the difference between the operational schedule and the actual consumption as stated in (6). The operational schedule is sent to the TSO the day before operation. As the actual baseline of the portfolio is unknown before operation, the aggregator will have to use the best available baseline prediction instead, i.e. $P_{\text{obs}}(t) := \hat{P}_{\text{base}}(t)$. If the baseline prediction equals the actual baseline we have no issues; however, if the actual baseline consumption deviates from the baseline prediction, the aggregator will have to use the portfolio’s energy capacity to compensate for the inaccurate operational schedule. Consequently, the energy capacity will be limited based on the accuracy of the baseline prediction.

The operational schedule is reported every day. Using the simple uncertainty model in (5) it is evident that over a day, the worst case energy delivery due to an uncertain energy prediction will be $\pm 24 \text{ hours} \cdot P^{\text{base}}$. As an example, consider a portfolio of 1,000 heat pumps with $P^{\text{base}} = 0.2 \text{ MW}$ (this inaccuracy is based on [40], see Sec. 8.3). Then the worst case situation for the portfolio is that the prediction error over the course of 24 h accumulates to 24 hours-$0.4 \text{ MW} = 9.6 \text{ MWh}$ which is more than the total energy capacity $E^{\text{cap}} = 8.0 \text{ MWh}$ of the portfolio of houses under consideration. Consequently, we cannot guarantee to follow the submitted operational schedule during the day and will thus not be able to participate in the ancillary service markets at all. This clearly illustrates how the uncertain baseline predictions can influence the possibilities to participate in the ancillary service markets.

6.3. Suboptimal market operation

One way to overcome the energy limitations of the flexible consumers is to provide ancillary services as combined deliveries,
where the portfolio of flexible consumption devices is combined with conventional generators. This can be done for example if a market player owns a portfolio of flexible devices with high ramping limits and low startup time and also owns a slower conventional generator. Depending on the devices’ properties, it may be possible for the market player to design a control strategy that allows the fast portfolio and the slow generator unit to collectively provide primary reserve. Hereby, the player can gain from the flexible consumers to increase the value of the slower generator, which else would only be able to participate in the less attractive markets for secondary or tertiary reserve.

However, now consider the case where a second player has a generator able to provide secondary reserve at a lower cost than the first player. Seen from a global perspective, it would be optimal if the cheaper secondary reserve generator of the second player was used together with the flexible consumer portfolio of the first player to provide a combined delivery. However, as these two devices are owned or operated by different players, such combined delivery cannot be handled under current market regulations. Consequently, suboptimal market operation can occur when players combine local devices to provide faster reserves. The method we propose in this paper exactly solves this issue by coupling the ancillary service markets.

7. Proposal of market interaction

In the previous section we have illustrated three major issues of using a portfolio of flexible consumption devices as providers of ancillary services. The first two issues deal with the energy limitation and the uncertain baseline. The third issue illustrates how combined deliveries can lead to suboptimal market operation.

In this work we propose the following approach to improve the possibility for flexible consumers to participate in the fast ancillary service markets.

Proposal. Operational schedules can be continuously adjusted throughout the delivery period. The adjustment must satisfy the ramping and latency constraints of secondary or tertiary control. If the operational schedule is adjusted according to the secondary control constraints, the cost of secondary control shall apply for the difference between the original operational schedule and the adjusted operational schedule; similarly, if the operational schedule is adjusted according to the tertiary control constraints, the costs of tertiary control shall apply.

Notice that although we propose a very specific method, this should merely be seen as an example. The main message of this paper is not this exact proposed method; rather, that we in general can increase the possibilities for flexible consumers to participate in the ancillary service markets by having well-defined regulations that allow continuous adjustments of the operational schedule at a well-defined cost.

7.1. Illustration of proposed method

The sequence diagram in Fig. 6 illustrates the proposed method in the market context. The first actor in the sequence diagram is the aggregator who utilizes a portfolio of flexible consumption devices in the ancillary service markets, and uses the proposed method to restore the portfolio energy level. The second actor is the TSO, who is the buyer of ancillary service, and finally, the third actor is the remaining providers of ancillary services.

As the figure illustrates, the aggregator will submit an operational schedule day-ahead describing the following day’s consumption with a given resolution according to current regulations (see also Sec. 4.2). Intra-day, the proposed method allows the aggregator to adjust the operational schedule continuously. This means that the aggregator at any time intra-day can submit an adjusted operational schedule according to the limitations described in Sec. 7; following, the TSO will confirm the adjusted operational schedule provided the adjustment satisfies the regulation. Next, the TSO will compensate for the adjustment of the operational schedule by activating the necessary reserves from the other ancillary service providers. One such an adjustment is illustrated in Fig. 6.

7.2. Comparison with existing regulations

Denmark is among the most active smart grid countries in Europe, therefore it is interesting to compare the method proposed above to the current Danish ancillary service regulations.

The Danish regulations describe that market players already now indeed are allowed to adjust a previously submitted operational schedule [19,41]. The regulations do, however, differ significantly from the method proposed in this work as elaborated in the following.

- The regulations specify that the operational schedule can be adjusted in case a difference between actual operation and the submitted operational schedule larger than 10 MW is detected. Consequently, the possibility to adjust the operational schedules is a way for the TSO to be aware of larger outages. Hence, it is different from the method proposed in this work where the operational schedule is adjusted in a continuous manner to restore the energy level of the flexible consumers. The proposal in this work is not only meant as a way for the TSO to be aware if an ancillary service provider has a larger outage; rather, we propose to deliberately couple the markets by allowing market players to actively and continuously adjust operational schedules. This allows the TSO to continuously utilize the slower reserve to compensate for operational schedule adjustments.

- The regulations do not state under what constraints the operational schedule can be adjusted, only that a latency time of 5 min must be honored. Consequently, a market player making a rapid change in the operational schedule can cause activation of faster reserves at no cost causing a loss for the TSO. This is therefore not a sustainable solution if a large number of market players will perform continuous operational schedule...
adjustments, as this potentially can generate a large economical deficit for the TSO. In the proposed method, the TSO covers its expenses by charging the aggregator according to the operational schedule adjustments.

Also in the ENTSO-E handbook [37,38], no specifications of operational schedule adjustments are mentioned.

8. Numerical results

In this section, we present a number of numerical results that illustrate the benefit of allowing continuous operational schedule adjustments according to the proposal in Sec. 7. First, we illustrate the overall concept; following, we illustrate how the method is able to handle both energy limitations and inaccurate baseline predictions.

8.1. Illustration of the overall concept

We illustrate the overall concept using the following example. Consider a portfolio of flexible consumers with parameters

\[
\begin{align*}
    &P_{\text{min}} = 0 \text{ MW}, \quad P_{\text{max}} = 2 \text{ MW}, \\
    &E_{\text{max}} = -E_{\text{min}} = 0.1 \text{ MWh}, \quad P_{\text{base}}(t) = 1 \text{ MW}, \\
    &P_{\text{os}}(t) = 0.5 \text{ MW}
\end{align*}
\]

(11)

i.e. the aggregator has offered a symmetric primary reserve equal to half of its capacity \(P_{\text{os}} = 0.5 \text{ MW}\). The aggregator has further submitted a constant operational schedule \(P_{\text{os}}(t) = P_{\text{base}}(t) = 1 \text{ MW}\) to the TSO. This could correspond to a portfolio of battery systems with a low energy capacity or a portfolio of thermal devices with very tight temperature bounds (see Table 2). The relatively small energy capacity is chosen deliberately to illustrate the presented method’s ability to use such devices in the ancillary service markets.

We consider the extreme power reference illustrated in subplot 1 of Fig. 7 (purple dashed line): after 5 min the reference changes from 0 MW to the maximum delivery of 0.5 MW and following, after 25 min, the reference changes to the other extreme of \(-0.5 \text{ MW}\). For primary reserve, the reference depends on the system frequency and we notice that the presented reference is highly unlikely; however, we have deliberately constructed it to illustrate the overall concept. Later, real life frequency measurements will be used to construct realistic power references.

Two scenarios are considered: a case where the aggregator is not allowed to adjust the operational schedule and a case where the aggregator is allowed to adjust the operational schedule as proposed in this work. Both cases are illustrated in Fig. 7 and show that the aggregator is able to track the power reference when allowed to adjust the operational schedule while it is not able to track the reference when adjustments are not allowed. Let us examine this further to gain insight in the presented method.

First, we examine the behavior of the conventional method where the operational schedule is not adjusted. At the first instance at time 5 min, the portfolio reduces its consumption to 0.5 MW as seen in subplot 2; however, the energy storage is empty after 12 min as shown in subplot 3 and the portfolio fails to track the reference as seen in subplot 1. The conventional method also fails at time 50 min, this time because the storage is full.

Now let us examine how the proposed operational schedule adjustment method works. First as the reference changes from 0 MW to 0.5 MW at time 5 min, the portfolio delivers the full response as evident from subplot 2, however, the aggregator starts adjusting the operational schedule as seen in subplot 4. The adjustments are made under the constraints of secondary reserve such that the TSO is able to activate secondary reserve accordingly as illustrated previously in Fig. 6. After the latency time of 30 s and the ramping time of 6 min, the aggregator has adjusted the operational schedule from 1 MW to 1.5 MW and restored the consumption to the nominal 1 MW as seen in subplot 2. Hereby, the reference of 0.5 MW is still tracked \(P_{\text{os}}(t) = 0.5 \text{ MW}\) as requested while the portfolio consumes the desired 1 MW; consequently, the energy storage does not saturate, as shown in subplot 3. As the power reference changes from 0.5 MW to \(-0.5 \text{ MW}\) at time 25 min, the portfolio must deliver the full change of 1 MW causing the consumption to be 2 MW; following, the operational schedule is adjusted such that the consumption can be restored to the desired 1 MW ensuring that the energy storage does not saturate. The figure further illustrates that in this worst case scenario, the aggregator will not be able to place higher bids than half of its capacity \(P_{\text{os}} = P_{\text{max}}/2\) as seen in subplot 2.

Finally, Fig. 7 can be used to determine the energy storage \(E_{\text{worst}}\) required to handle this worst case situation. The worst case energy is given by

\[
E_{\text{worst}} = 2P_{\text{os}}(t_{2}^{\text{ramp}} + t_{2}^{\text{lat}} - t_{1}^{\text{ramp}})
\]

(12)

as this is the required energy deliver if the reference changes from one extreme to the other. In Fig. 7 subplot 2 \(E_{\text{worst}}\), corresponds to the area between the baseline of 1 MW and the triangular shaped power consumption at the time of the worst case situation at time 25 min until it is restored at time 37 min.

The correlation (12) can be used to determine how much power an aggregator at most is able to bid into the primary reserve market.
Again we consider the energy storage described above with $E^{\text{cap}} = 1 \, \text{MW}$. The curve in Fig. 8 shows the maximum power bid we are able to place in the primary reserve market depending on the available energy capacity $E^{\text{cap}}$ by utilizing the secondary or tertiary reserve to restore the storage energy level. As the figure shows, using the secondary reserve makes it possible to offer a reserve of $P^{\text{sec}} = P^{\text{cap}}/2 = 0.5 \, \text{MW}$ with an energy capacity $E^{\text{cap}} \geq 0.11 \, \text{MWh}$ while an energy capacity $E^{\text{cap}} \geq 0.25 \, \text{MWh}$ is required when relying on the slower tertiary reserve to restore the energy storage. This is clearly much more favorable for the aggregator compared to a worst case energy capacity requirement of 84 MWh if we are not allowed to adjust the operational schedule, see Sec. 6.1. This reveals how the presented method allows an aggregator to participate in the primary reserve market on much better terms.

### 8.2. Simulation example I: limited energy storage

In this subsection we consider how the proposed method resolves the first issue, namely the limited energy storage. We consider a portfolio with the same parameters as the previous example, see (11), i.e., we have a portfolio of flexible consumption device with a constant baseline consumption 1 MW which it is able to vary around with ±1 MW however under strict energy limitations of 0.1 MWh.

Based on the worst case consideration presented in (12) it can be seen that by relying on the secondary reserve, we are able to provide $P^{\text{sec}} = 0.5 \, \text{MW}$ of primary reserve. By comparison, a worst case situation without operational schedule adjustments would limit the bid to $P^{\text{sec}} = 0.1 / (2-168) \, \text{MW} = 0.0003 \, \text{MW}$ which in practice means this device would not be suitable for primary reserve. Again this shows the benefit of the presented method.

Now we use real frequency measurements from the ENTSO-E grid to compare the proposed method where we continuously adjust the operational schedule to a conventional situation where the operational schedule is not adjusted. We do this via simulations based on the model presented in Sec. 4.2. The simulation is conducted as follows. The historical grid frequency deviation measurements $\Delta f$ is translated to a certain required power consumption for the portfolio according to the ENTSO-E specifications for primary frequency control, see Sec. 5.2. The sampling time is 1 s, as required by the regulations. For the conventional case, we simply let the portfolio consume the required electricity according to the reference dictated by the grid frequency deviations and examine the resulting energy level. This benchmark case is then compared to a case where the proposed method is utilized to restore the energy level via operational schedule adjustments. In this simulation, a simple controller is implemented that seeks to restore the portfolio energy level by continuously adjusting the operational schedule. This is further made clear in the following concrete simulation results.

In Fig. 9, a 4-h period of operation is illustrated based on the real life frequency measurements presented in subplot 1. Subplot 2 shows the resulting power consumption of the portfolio in the two situations illustrating that both strategies provide fast responses according to the demand. The consumption of the conventional strategy is directly dictated by the grid frequency deviation $\Delta f$; on the contrary, the consumption in the case where operational schedule adjustments are allowed is a function both of the grid frequency deviation but also of how the operational schedule is adjusted, see Fig. 4. Subplot 3 shows the energy level of the portfolio. This plot reveals that the conventional method with no operational schedule adjustments will require an energy delivery that is far outside the limits of the portfolio, while the presented method is able to stay within the limits. Subplot 4 shows the fixed operational schedule compared to the adjusted operational schedule. The operational schedule is adjusted under the latency and ramping constraints of secondary reserve which is the reason for the low frequency content in this signal.

The figure shows the important result that the continuous operational schedule adjustment method allows an energy limited portfolio of flexible consumption devices to utilize its strength of being able to provide fast regulation without being driven away from the energy midpoint.

To further investigate the method, we perform a number of 1-week long simulations using real ENTSO-E frequency measurements. The simulations show that the operational schedule

![Fig. 8. The maximum power we are able to bid into the primary reserve depending on the energy capacity $E^{\text{cap}}$ when utilizing either the secondary or tertiary reserve to restore the energy level.](image-url)
adjustments during the course of a week sum up to around 6 MWh of both upward and downward regulation when providing a symmetric delivery of $P_{res}^{acc} = 0.5$ MW. The adjustments are made according to the constraints of secondary reserve and consequently the price of secondary reserve applies. The price is $\pi_{2} \approx 13\€/MWh$ in the Western Danish market which yields an expense around 150€ for a week. By comparison, the income per symmetric MW of primary reserve capacity is $\pi_{1} \approx 30\€/MWh$ yielding a total income in the order of 2,500€ for the delivery of $P_{res}^{acc} = 0.5$ MW for one full week. This illustrates that the fast regulation of the flexible consumption devices is very valuable compared to the small amounts of secondary reserve that must be purchased to continuously restore the energy level. In other words: the method allows the flexible consumers to deliver the expensive fast reserve while the inexpensive slow responses are shifted to the slower reserve providers.

By simulating over several weeks we further discover that when the contracted delivery is $P_{res}^{acc} = 0.5$ MW, the highest consumption of the portfolio is in the order of 1.3 MW while the smallest consumption is in the order of 0.7 MW which is very conservative compared to the limits $P_{max} = 2$ MW and $P_{min} = 0$ MW. The reason is that we have dimensioned the bid $P_{res}^{acc}$ after the worst case situation as described in Sec. 8.3; however, these simulations indicate that it might be possible to find a way to be less conservative such that bids close to the total capacity can be made, i.e. that we in this example would be able to offer $P_{res}^{acc} = 1$ MW. This study is, however, outside the scope of this work.

8.3. Simulation example II: uncertain baseline

In this second example we illustrate how the proposed method resolves the issue of uncertain baseline predictions. We consider the real life case presented in Ref. [40] where the baseline consumption of a portfolio of heat pumps is examined. The uncertainty arises from the fact that the outdoor temperature, the solar irradiation, and the human behavior cannot be predicted accurately. In Fig. 10 we show the real life power consumption of a heat pump portfolio along with a prediction of the consumption made the day before. The results are taken from Ref. [40] and scaled from a portfolio of 40 heat pumps to a portfolio of 1,000 heat pumps revealing an inaccuracy in the order of $P_{acc} = 0.2$ MW for the entire portfolio. The energy parameter is set to $E^{sp} = 8$ MWh for the portfolio (see Table 2).

We compare the two strategies: continuous baseline adjustments versus no baseline adjustments for the course of one week. The contracted reserve is $P_{res}^{acc} = P_{cap} / 2 = 0.3$ MW. To clearly show the effect of uncertain baseline predictions, we assume that no energy delivery is required during the week; consequently, the heat pump portfolio must simply assure that no power delivery is made, which corresponds to tracking the submitted operational schedule. The second subplot of Fig. 10 clearly shows that if the operational schedule is not adjusted, the energy limitations will be violated due to the inaccurate baseline predictions; however, by allowing the operational schedule to be adjusted the energy level can be kept close to the energy midpoint.

Again, we consider the economic aspects. Over the course of a week, the amount of purchased upward and downward secondary reserve is each in the order of 3 MWh yielding a total expense of 76€ whereas the value of a symmetric primary reserve delivery of 0.3 MW is in the order of 1,500€. This clearly shows that the presented method is able to let the portfolio deliver the valuable fast responses while the slower and cheaper responses are shifted to the providers of secondary reserve.

9. Conclusion

In this paper we considered an aggregator that provided ancillary services based on a portfolio of flexible consumption devices. We proposed a method where the aggregator was allowed to continuously adjust its operational schedule and thereby restore the energy level of the flexible consumption devices. This made it possible to utilize flexible consumption devices with energy limitations and with inaccurate baseline predictions to participate in the ancillary service markets to a much larger extent than under the current regulations.

The proposed method was illustrated through two numerical examples, one example where an aggregator of flexible consumption devices was characterized with a very low energy capacity, and another example where only an inaccurate consumption baseline was available. In both examples, the proposed method was able to radically increase the feasible reserve bids compared to a situation where the aggregator was not allowed to adjust the operational schedule.

Consequently, the method proposed in this work allows new providers of fast ancillary services to be able to enter the electricity markets and possibly replace the conventional fossil fuel based ancillary service providers.

Acknowledgments

The work is completed as a part the iPower project which is supported by the Danish government via the DSR-SPIR program 10-095378.

References
