A Market Design to integrate Demand Response into NEM Pricing and Dispatch

(previously referred to as “Scheduled Lite”)

Prepared by

Creative Energy Consulting

July 2024

Executive Summary

Background and Context

Under the current arrangements, resources connected to the NEM power system - generators, loads and storage schemes - are either scheduled or non-scheduled. AEMO has visibility and control of the former but not the latter. This has not mattered historically, because the behaviour of the non-scheduled resources – typically assets owned by customers and served by retailers – has been relatively stable and predictable. AEMO has been able to forecast the power demand created by unscheduled resources and dispatch scheduled resources to meet it. This dispatch process, carried out every five minutes, has maintained a good balance between supply and demand across the NEM.

But ongoing growth in new flexible resources (like batteries and EVs), new automated control mechanisms, and new retailer business models are likely to drive a significant change in behaviour in the form of responsiveness of these resources to the NEM spot prices set by the dispatch process. In short, it is increasingly the case that, other things being equal, if spot prices go up, demand goes down, and vice versa.

Without a change to the NEM design, this increasing demand response (DR), not visible to AEMO in the dispatch process, will lead to growing supply-demand imbalances in dispatch. This will lead to increasing amounts and costs of the frequency regulation needed to correct these imbalances. DR will also lead to corresponding price-driven variability in metered demand which is liable to adversely affect AEMO's demand forecasting accuracy, potentially exacerbating these imbalances.

The AEMC engaged Creative Energy Consulting to develop a market design that can help manage these imbalances, by encouraging retailers to make their demand response - and that of their customers – more visible to AEMO, who can then integrate it into dispatch. The proposed design is described in this report and summarised in figure E1 below.

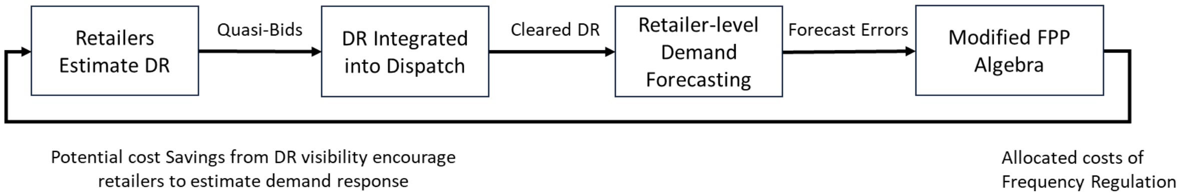

Figure E1: Overview of Proposed Design

The proposed design is described and discussed below.

Integrating Demand Response into Dispatch

Imbalances can be managed by integrating DR into the dispatch process, as shown in figure E2 below. Under the current dispatch process, with DR invisible to AEMO, the demand forecast used by AEMO in each dispatch is a single number, independent of spot price, represented by the vertical purple line in the figure below. Scheduled resources submit offers to generate, represented by the blue line. The dispatch process essentially finds where these supply and demand curves intersect and sets the spot price accordingly. It also dispatches sufficient generation to supply this forecast demand.

But, with growing DR, the actual demand curve is not vertical but downward sloping: the higher the spot price, the lower the demand, represented by the green line in the figure. When the spot price is high, actual demand could be much less than AEMO's forecast. So DR will lead to imbalances between supply and demand, especially at very high and low prices.

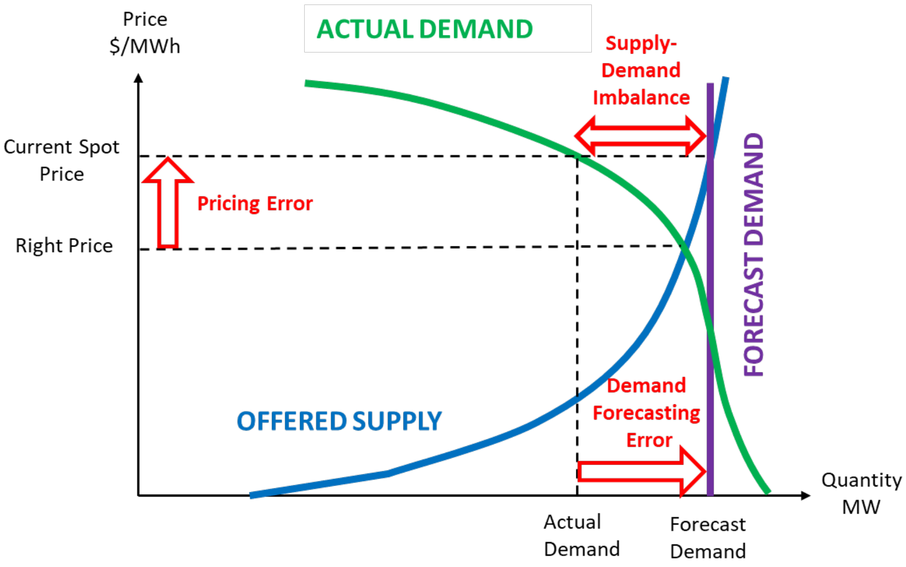

Figure E2: Integrating DR into Dispatch

The demand-supply imbalance is caused by DR, which is the difference between the purple and green lines. Integrating demand response into dispatch, means using the green line rather than the purple line to represent demand. So, the spot price is now set according to the intersection of these two lines, and generation is dispatched to supply the corresponding demand quantity. There is now no forecasting error and so no supply-demand imbalance. The spot price is now set at the "right price"; the price at which demand and supply are in balance.

But for this integration to be possible, demand response must be known to dispatch, so that the green line can be constructed. In the proposed design, retailers to each estimate their own DR and submit these estimates to AEMO in the form of "quasi-bids". This approach is considered likely to yield more accurate DR estimates from retailers than AEMO making its own estimates. Retailers know their customers, and the signals they are sending to them, and so can build estimates from the bottom up; AEMO would have to use a top-down methodology, which would likely be less accurate and also more contentious, particularly given that these DR estimates would – by design – significantly impact spot price outcomes.

Retailer estimates of DR that are incorporated into the dispatch process, becomes "visible", in contrast to the "invisible" DR in the current market. The design proposes that retailers should be financially incentivised to make their DR visible, rather than this being mandated. Furthermore, the incentives should be predicated on the accuracy of the DR estimates, so that retailers put additional resources into improving accuracy where this can be done at a reasonable and proportionate cost.

DR visibility not only improves dispatch but also allows AEMO to improve its demand forecasting by correcting the demand actuals to remove spot price impacts. AEMO will continue to forecast this price-corrected demand, referred to here as "base demand[1]". This is preferred to the alternative of assigning this role to retailers. Retailers do not have the real-time data that AEMO has, and so could not forecast base demand with AEMO's accuracy.

Imbalance Costs and Frequency Performance Payments

Imbalances between supply and demand in real-time quickly lead to power system frequency deviating from its target 50Hz level, and frequency regulation is required to keep frequency within operational limits. This regulation is provided by scheduled resources adjusting their supply, effectively correcting any imbalances. The cost of this frequency regulation is shared between generators and retailers; if imbalances grow as a result of invisible DR, these costs will grow too.

The philosophy of the proposed design is to offer rewards to retailers who make their DR visible and to set these rewards at a level that reflects the value of this visibility. Since the value comes from reducing imbalances and so reducing the costs of frequency regulation, the rewards should be based on these savings. However, estimating the nature and magnitude of these savings is complicated.

Under a recent rule change, due to be implemented in 2025, frequency regulation costs are to be estimated and allocated using a new frequency performance payments (FPP) mechanism. For generators, the FPP scheme uses 4-second (4S) generator metering to precisely calculate the amounts by which generators deviate from their dispatch target. It then uses a statistical analysis to infer which generators are exacerbating imbalances and which are helping to correct them, and then either charges or pays these generators, respectively.

Customers generally don't have 4S metering, so retailer deviations and their associated costs are not known directly; they can only be inferred from the available 5-minute metering data. The FPP scheme doesn't attempt to do this, but simply allocates costs in proportion to retailer size. This is a reasonable approximation if no retailers have DR; but, where some do, it fails to distinguish between retailers with visible DR, those with invisible DR, and those with neither. The proposed design adapts and enhances the FPP scheme to allow these distinctions to be made and for costs to be allocated accordingly, as explained below.

Identifying Demand Deviations arising from Demand Response

The starting point in making these distinctions is the insight that invisible DR will lead to larger demand forecasting errors, as noted above. In the current design, demand is only forecast at the regional level and so forecast errors cannot be attributed to individual retailers. The proposed design therefore introduces a new retailer-level demand forecasting process to forecast retailer demand and calculate associated forecast errors. This must be done ex-post, as part of the settlement process, because this is when the necessary customer metering information becomes available, which can then be aggregated for each retailer. The new process will use existing AEMO dispatch forecasting models, applied at a retailer - rather than regional - level.

Retailers with and without DR will have different kinds of deviations. Those without DR will have demand that simply varies randomly around the forecast, and deviations will reflect this "noise". DR, however, is not at all random but rather prompted by changes in the spot price. So DR will create

---

[1] So the purple line in figure E2 becomes the forecast *base* demand, rather than just the forecast demand.

step changes in retailer demand at the start of each new dispatch interval as the spot price changes and the quantity of DR adjusts accordingly.

Based on these insights and inferences, the “demand deviation” calculated by the FPP scheme, and attributed to retailers in aggregate, can be decomposed into three deviation components, attributable to visible DR, invisible DR and underlying demand variability. Frequency regulation costs associated with each of these components can then be calculated by applying the existing FPP algebra separately to each of these three deviation components. So the aggregate cost charged to retailers in the current FPP scheme has now been divided into three parts.

Allocating the Costs of Deviations

The final stage is to allocate costs between retailers. Three different allocation metrics are used, reflecting the driver of each respective deviation and cost component.

Firstly, the costs caused by visible DR are allocated between retailers in proportion to the change in cleared DR for a retailer between consecutive dispatch intervals. Whilst the amount of visible DR is known and incorporated into dispatch, the change in DR nevertheless creates step changes in demand that impact on frequency regulation.

Secondly, the costs caused by invisible DR are allocated between retailers in proportion to retailer demand forecasting errors. The rationale here is that, though the exact amount of invisible DR is unknown, by definition, it can be inferred by the size of the forecasting errors, based on the insight that invisible DR leads to larger forecasting errors.

Thirdly, the costs arising from underlying demand variability are allocated in proportion to retailer size: ie retail demand. This is similar to the existing FPP design, and reflects the fact that, in the absence of DR, retailer demand variability generally impacts on frequency regulation costs in proportion to retailer size.

Under this revised FPP scheme, retailers with invisible DR will make a larger contribution to frequency regulation costs, other things being equal. They can reduce this contribution by making their DR visible: that is, by estimating the response and submitting these estimates in bids to AEMO, who can then incorporate them into the dispatch process. That, in turn, will lead to a reduction in frequency regulation costs. So the desired incentives for DR visibility have been established.

Staged Implementation

The proposed design could be implemented in two stages, as follows. Initially, the new settlement algebra would be implemented by AEMO in a “shadow mode” where the new amounts are calculated and published, but actual financial transactions continue to be based on existing rules. Only once material differences between the outcomes of the two settlement methods are seen would the proposed design “go live”, with the new bidding, dispatch and settlement processes all operational. Retailers would then individually decide whether and how to estimate and bid their DR.

Assessment of Costs and Benefits

The proposed design involves four new or modified market processes:

1. *DR quasi bidding*: the estimation of DR by retailers and submission of these estimates to AEMO;
2. *Dispatch*: the inclusion of these estimates into AEMO's dispatch and demand forecasting processes;
3. *Retailer-level demand forecasting*: a new process, carried out ex-post, and used to identify retailers that have invisible DR based on this leading to higher forecasting errors; and
4. *Modified FPP*: a revised FPP scheme that decomposes the aggregate demand deviation and then allocates the associated frequency regulation costs between retailers using different allocation metrics.

Costs from the first process are incurred by those retailers who opt to participate. They will do so only if the financial incentives – reflecting the savings in frequency regulation costs – exceed these costs. Therefore the net benefit, from lower frequency regulation costs but higher retailer costs, should always be positive: at worst, if DR is immaterial then retailers simply won't engage, and no new retailer costs are incurred.

The other three processes will be developed and operated by AEMO, who will incur costs whether retailers participate or not. By largely adapting existing market processes rather than creating brand new ones, these AEMO costs should be kept modest. They can be compared against the broader benefits from implementing the proposed design, in lower dispatch costs, and more stable and predictable spot prices.

This broader benefit of DR visibility is expected to grow over time, as DR develops and expands. The staged implementation approach means that only the costs of shadow operation are incurred initially, with the full cost incurred only once "go live" is triggered by material differences between the shadow and actual settlement amounts. At that point, invisible DR will be materially impacting on dispatch efficiency and so DR visibility will become valuable and necessary.

Questions and Answers

The table below summarises answers to some questions that might arise in relation to the proposed design, around its rationale, performance and scope. These issues are discussed in detail in this report.

Question	Answer
Does the proposed design only affect retailers	It also affects small generation aggregators and, generally, any market participant financially responsible for non-scheduled resources.
Can AEMO's existing regional demand forecasting methods be repurposed	Conceptually, yes. AEMO is best placed to advise on practical and operational issues.
Are all responses of demand to price regarded as demand response?	All responses that AEMO does not model are considered DR.
Can retailers accurately estimate demand response?	Accuracy will not be perfect but will be fit-for-purpose. Retailers will likely specialise.
Can retailers forecast their base demand too?	No. Not without real-time metering data. This is why this is left to AEMO, who has this data.
Should visible DR be charged for associated frequency regulation costs?	Yes. This may incentivise visible DR which helps restore system balance following a generator outage
Should visible DR enjoy lower regulation FCAS costs?	Maybe. It would improve incentives, but could lead to higher FCAS costs for retailers overall.
How are retailers with 4S metering managed?	Unclear. It may be possible to adapt the design to make use of this data and provide better incentives.
How does distribution congestion affect the proposed design?	DR could be scaled back by DNSPs. Retailers should factor that into their DR bids.
How does transmission congestion affect the proposed design?	DR would be bid regionally and cannot be constrained to help manage congestion.

Table of Contents

1. Introduction and Background

AEMO Rule Change Proposal

In January 2023, AEMO submitted a *scheduled lite* rule change proposal, designed to improve demand forecasting and generation dispatch in the context of growing customer exposure and responsiveness to NEM spot prices. The objective is to ensure this growing responsiveness is appropriately accommodated and reflected in these core market processes, ensuring the maintenance and continuation of efficient and secure dispatch.

Without reform, there is a concern that this growing response could lead to larger demand forecasting errors, which in turn will require more frequency regulation to correct the resulting supply-demand imbalances. This would lead to higher costs, borne by generators and customers.

Our Engagement

The AEMC engaged Creative Energy Consulting to develop an alternative scheduled lite design. This has similar objectives to the AEMO proposal but some key differences in philosophy and design. It is founded on a set of adopted principles that are explained in chapter 3.

This report describes and discusses the proposed design.

In this report,

- *current design* refers to the existing Rules;
- *proposed design* refers to our proposed design as set out in this paper.

CEC was originally engaged by the AEMC in October 2023, and a version of this report titled *Scheduled Lite design to integrate Demand Response into NEM Pricing and Dispatch* was prepared and published by the AEMC in December 2023. CEC was re-engaged by the AEMC in February 2024 to provide further advice on the proposed design and to assist with engagement with stakeholders on it. As a result, some changes have been made to the proposed design and these changes are reflected in this second version of the report. The key changes are listed in Appendix B.

Report Structure

This report is structured as follows:

- Chapter 2 describes how and why growing price-response is expected, in the absence of rule changes, to lead to increasing demand forecasting errors and higher costs of frequency regulation;
- Chapter 3 provides an overview of the proposed design, listing the key principles it is built upon and outlining its philosophy and architecture;
- Chapters 4-7 describe the key elements of the proposed design in more detail, their underlying rationale and expected outcomes;
- Chapter 8 aims to anticipate and answer possible stakeholder questions around the design rationale, details and possible extensions or amendments;
- Chapter 9 describes a possible implementation strategy; and
- Chapter 10 qualitatively assesses likely costs and benefits.

There are two appendices. Appendix A describes some possible retailer business models that would give rise to demand response. Appendix B describes changes between the 2023 and 2024 versions of this report.

Glossary of Terms

To improve readability, this report endeavours to use plain English, meaning that some of the terminology used may, at face value, be ambiguous or approximate. Nevertheless, the terms used, as listed in table 1 below, have a strict meaning in terms of existing NEM concepts. Acronyms used are shown in table 2 below.

Term	Meaning
Base Demand	The anchor around which demand response is defined; the demand level at the base price
Base Price	The spot price level, chosen or implied, at which demand is equal to the base demand: ie by definition there is zero demand response at this price
Cleared DR	The amount of demand response that is bid at the outturn spot price
Demand	Net non-scheduled load: that is the aggregate consumption of non-scheduled loads minus the aggregate output of non-scheduled generators. This includes any charging or discharging of non-scheduled storage schemes. Because of demand response, demand is a function of spot price (as well as time), rather than a single number.
Demand Response (DR)	A change in demand as a direct or indirect response to a change in the spot price, or other short-term changes (eg dynamic network tariffs) not modelled in AEMO's demand forecasting algorithms. It is the difference between the base demand and actual demand. Demand response will generally be positive at high spot prices and can be negative at low spot prices
Generator	A NEM participant that is financially responsible for a scheduled resource: this could be a scheduled generator, semi-scheduled generator, scheduled load or scheduled integrated resource provider.
Quasi-bid	The submission of DR estimates to AEMO by retailers
Resource	A generator, load or storage scheme connected to the NEM power system
Retailer	A NEM participant that is financially responsible for a non-scheduled resource. In NEM terminology this could be retailer, a small generator aggregator or non-scheduled integrated resource provider.
Supply	Net scheduled generation: that is the aggregate output of scheduled and semi-scheduled generators minus the aggregate consumption of scheduled loads. This includes any charging or discharging of scheduled storage schemes. Supply is a function of spot price: ie the higher the spot price, the more generation dispatched, other things being equal.

Table 1: glossary of define terms used in this report

Acronym	**Meaning**
4S	Four-second [metering]
5M	Five-minute [metering]
BDU	Bi-directional Unit
DI	Dispatch Interval
DR	Demand response
FCAS$	Regulation FCAS cost allocation
FPP	Frequency Performance Payments
FPP$	FPP settlement amount
NEMDE	NEM Dispatch Engine
PASA	Projected Assessment of System Adequacy
PD	Pre-dispatch
RLF	Retailer-level Forecasting
SSG	Semi-scheduled generator
VPP	Virtual power plant

Table 2: Acronyms used in this report

2. Context and Objectives

Overview

AEMO’s dispatch process runs every five minutes and calculates dispatch quantities and spot prices. Whilst the former are critical in ensuring grid security, the latter are also important: not just in providing long-term signals to guide generation and retail tariffs but also in the short-term signals provided to some non-scheduled resources to adjust consumption or generation accordingly.

The latter will grow in significance as such *demand response* (DR) grows. Spot pricing should reflect this response, but cannot do so whilst this response remains invisible to the dispatch process. This will lead to growing imbalances between supply and demand, with increasing direct costs to provide the frequency regulation services needed to manage and correct these imbalances, as well as the indirect costs associated with spot prices not properly reflecting market conditions.

Retailers can today make this response *visible* through bidding scheduled load, but have no incentive to do so currently, because this does not affect the price or amount that they pay. Therefore the challenge, which the proposed design takes on, is to create new incentives that promote DR visibility, and to incorporate this visible DR into the dispatch process.

The NEM Auction

The dispatch process, sitting at the heart of the NEM, is an auction running every five-minutes. Like every auction, its outcomes contain two core elements:

- *Cleared quantities*: the amounts cleared in the auction.
- *Clearing prices*: the price or prices at which these amounts are traded.

In NEM dispatch, the focus has been on getting the *quantities* right, with the prices seen as secondary. That reflects the relative risks of getting it wrong. The cleared quantities must represent a secure dispatch, and an insecure dispatch can potentially lead to load shedding or even grid failure. The clearing prices – or spot prices – are important, but primarily through their influence on forward prices, retail tariffs and future investment. Wrong prices are unlikely to have implications for system security[2].

What is a “wrong” clearing price? The right price matches supply and demand; if the price is wrong the two will not match and there will either be a supply surplus or deficit, as shown in figure 1, below.

---

[2] indeed, spot prices are “wrong” in large areas of the NEM for large periods of time, due to transmission congestion that is not reflected in spot prices.

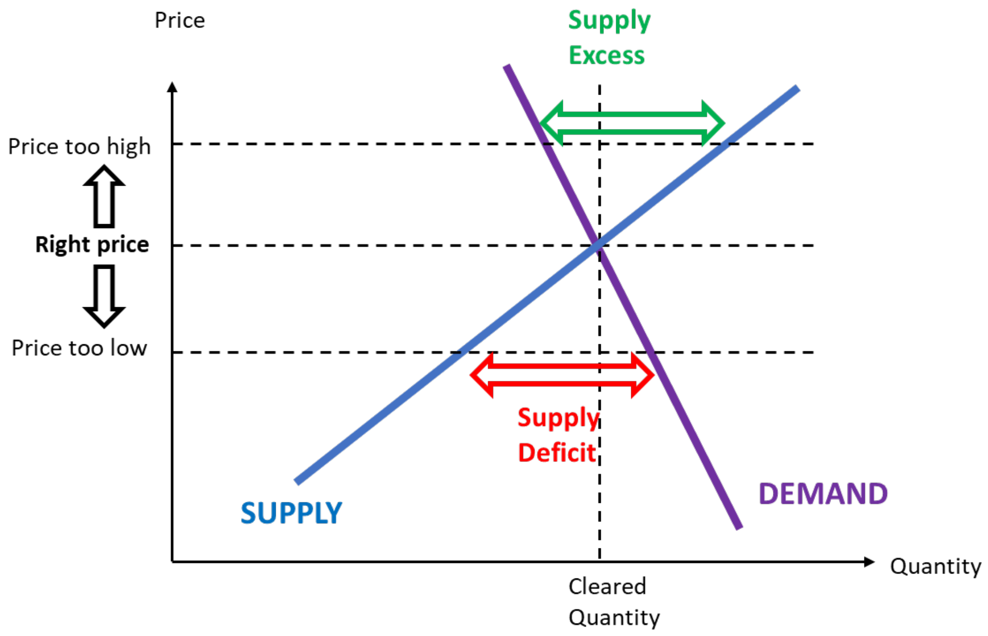

Figure 1: The right price ensures supply and demand are in balance

Supply and Demand in the NEM

In an electricity market, supply and demand could naturally be defined as generation and load, respectively. However, a more useful dichotomy, and the one used in this report[3], is based on whether or not AEMO is in control of the quantity produced or consumed. In the NEM, controlled quantities are referred to as scheduled[4]. Since any resource can either be a generator (producing electricity) or a load (consuming electricity), resources fall into 4 categories[5], as shown in table 3, below:

	Scheduled	**Non-scheduled**
Generation	*Scheduled Generation* Large generators Large storage schemes discharging	*Non-scheduled Generation* Small generators Rooftop PV Small batteries discharging
Load	*Scheduled Load* Large load opting to be scheduled Large storage schemes charging	*Non-scheduled Load* Customer loads Small batteries charging

Table 3: Categorisation of NEM Resources

---

[3] and also used by AEMO, who has a concept of "Operational Demand" which is similar (although for technical reasons not quite identical) to the demand concept used here.

[4] Generators can also be semi-scheduled, in which case AEMO is sometimes in control and sometimes not. For simplicity, these will be included as "scheduled" unless the context indicates otherwise.

[5] For reasons of simplicity, the terms in italics in the table do not precisely align with the NEM definitions

The following definitions are used in this report:

- Supply is scheduled generation minus scheduled load[6]; and
- Demand is non-scheduled demand minus non-scheduled generation.

In short, under this definition, supply is controlled by AEMO but demand is not. To maintain a balance between supply and demand, AEMO must forecast demand and then dispatch supply equal to this forecast.

In the NEM, every resource has an associated market participant that is financially responsible for it: ie financially settling with AEMO the dollar amounts associated with the resource. Whilst these market participants fall into several NEM categories, for simplicity this report refers to:

- *Generators*: being those financially responsible for scheduled resources; and
- *Retailers*: being those financially responsible for non-scheduled resources.

The extent to which prices are wrong, and the impact of these pricing errors on market imbalances, depends upon the price elasticity of demand: the higher the elasticity, the greater the pricing errors (for reasons discussed below) and the larger their impact. Supply elasticity is less important, because scheduled parties are required to operate at their cleared quantities (ie dispatch targets), irrespective of their preferences.

Demand is currently inelastic to spot price for several reasons:

- Few electricity consumers actually face spot prices, but rather are charged fixed tariffs, because these are simpler and have lower transaction costs;
- Whilst retailers *do* face spot prices, they are not permitted to control or instruct load without customer consent, which has to date been difficult to obtain;
- Responding to spot prices by controlling loads is typically costly and inconvenient compared to the benefits generated; and
- In any case, electricity demand typically has low short-term price elasticity for most end-uses.

It is useful to list these drivers, because there are expectations of these changing over the medium-term as:

- New or expanded end-uses (particularly batteries, EVs and electric water and space heating) can have relatively high short-term price elasticity, with such loads able to be shifted in time to avoid high prices;
- Automation of load control reduces the costs of active control; and
- Increasing spot price variation and volatility – due to increased penetration of variable renewables – increases the potential value of responding to spot prices.

---

[6] plus the new category of scheduled integrated resource providers, which combine generation and load, so could mean net generation or net load.

The demand side is currently represented in dispatch, as a fixed quantity, independent of price, with the quantity set by AEMO based on *short-term* (ie five-minute ahead) forecasting of demand. If demand is inelastic, then this vertical demand curve will be a reasonably accurate representation of current reality. As shown in figure 2, below, this approach might lead to some pricing errors but only minor imbalances.

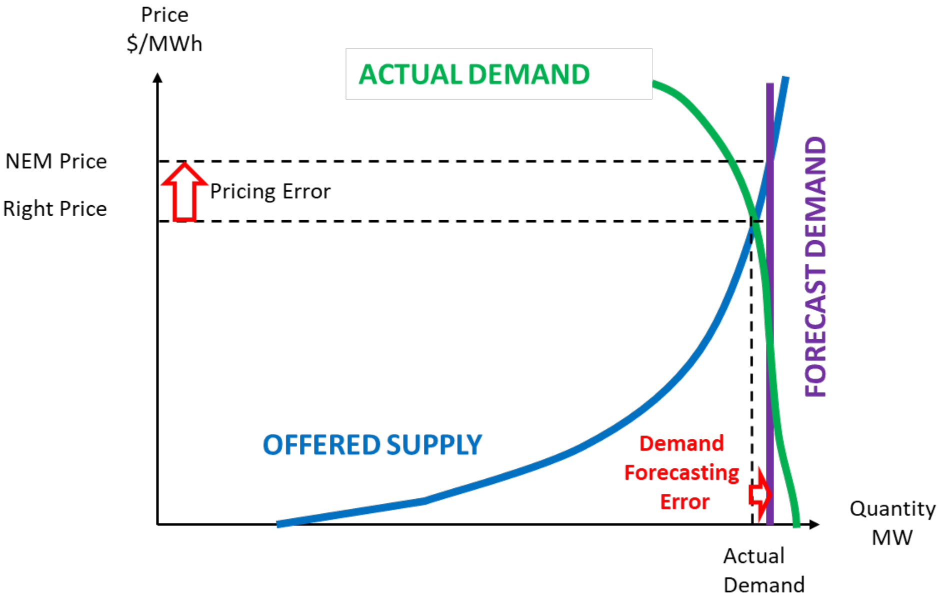

Figure 2: with inelastic demand, price and demand errors are small today

As demand response grows, pricing errors and imbalances will grow correspondingly, as shown in figure 3, below.

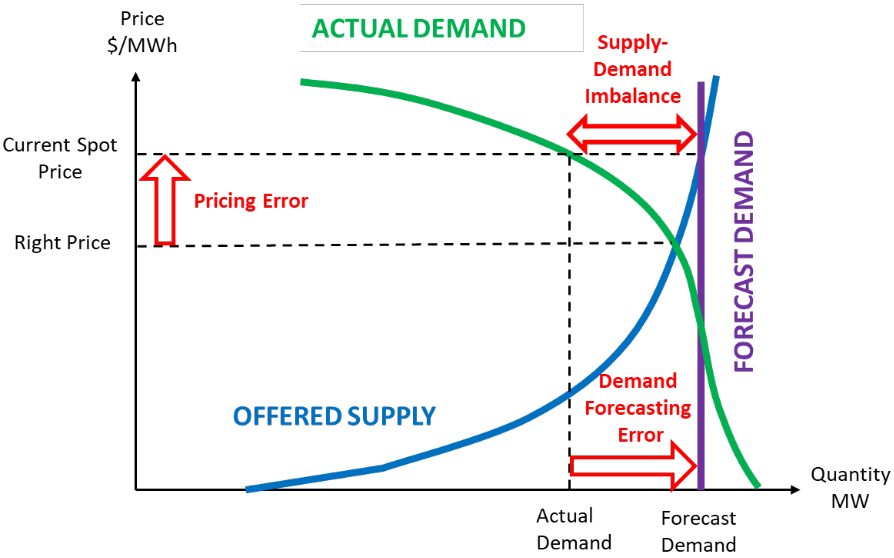

Figure 3: with elastic demand, price and demand errors would be much larger

Decomposing the Demand Curve

It is convenient to decompose the demand curve into:

- *base demand*: that is the demand level at some chosen *base price* point; and
- *demand response*: the difference between base demand and actual demand[7].

This is illustrated in figure 4, below.

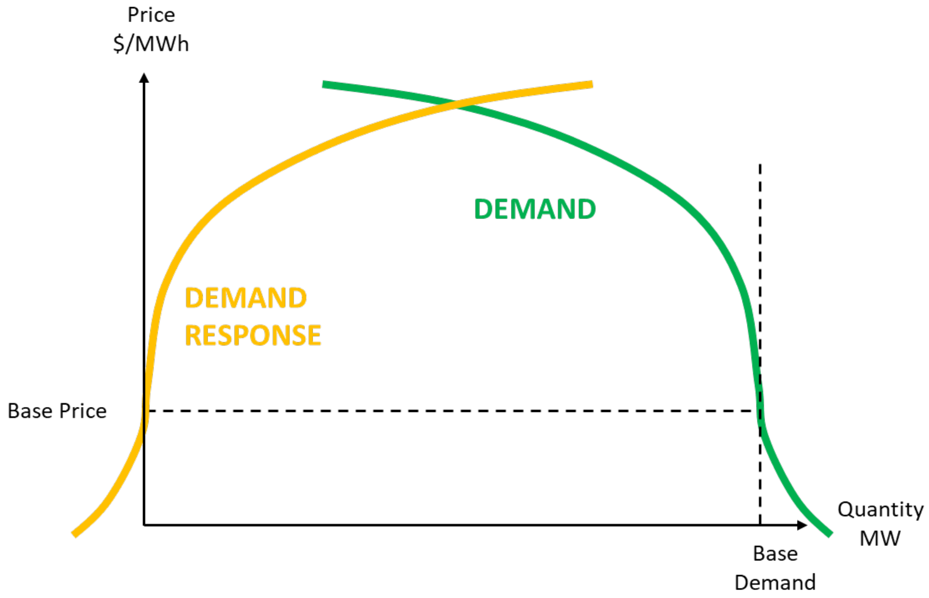

Figure 4: decomposition of demand into base demand and demand response

With demand being inelastic currently, AEMO is essentially just forecasting base demand. Demand response is small and AEMO does not attempt to estimate it or incorporate it into dispatch. However, if demand response were significant and its characteristics known, it could be incorporated into pricing and dispatch by finding the intersection of the supply and demand curves, as shown in the figures above.

Table 4 below presents a quantitative example of this approach.

---

[7] demand response will be positive where high spot prices lead to actual demand being lower than base demand; it may be negative when low spot prices lead to actual demand being higher than base demand

Price ($/MWh)	*Offered Supply (GW)*	*Base demand forecast (GW)*	*Demand Response (GW)*	*Demand minus Supply (GW)*
0	10	20	-1	11
5	13	20	0	7
10	15	20	0	5
50	17	20	0	3
100	18	20	0.1	1.9
500	19	20	0.2	0.8
1000	**19.5**	**20**	**0.5**	**0**
5000	**20**	**20**	**1**	**-1**
10000	20.5	20	3	-3.5

Dispatch outcome with DR included
Dispatch outcome with DR excluded

Table 4: example supply and demand curves

In today's design, the spot price would be set at $5000/MWh, being the price at which 20GW of supply is offered to dispatch, covering the forecast base demand. This would lead to 1GW of demand response, not included in the demand forecast, meaning an excess supply of 1GW: 20GW of dispatched generation but only 19GW of actual demand.

If, instead, the DR was included in dispatch, the spot price would be set at $1000/MWh: at that price, there is 19.5GW of demand, the 20GW of base demand minus the 0.5GW of estimated DR. There is also 19.5GW of supply offered at this price, so supply and demand are in exact balance.

DR is not known inherently and must be estimated. So who will estimate DR, and how and why they would do it? The proposed design creates financial incentives for retailers to estimate their DR and submit these estimates to AEMO to be incorporated into dispatch. Whether and how retailers do this is left to them.

Imbalances and Frequency Regulation

As discussed, wrong spot prices lead to supply-demand imbalances. To be precise, the imbalance is between *dispatched* supply and *actual* demand; *actual* supply and actual demand must be in exact balance at all times or the grid collapses. Dispatch cannot be changed; targets are locked in for the dispatch interval (DI) and only reviewed and revised at the start of the next DI. So imbalances must be managed through deviations away from dispatch targets. These will be prompted by the deviations in frequency that supply-demand imbalances cause. Supply must respond to *regulate* frequency and this, in turn, will ensure that supply and demand remain in balance.

There are three primary sources of this frequency regulation, as shown in table 5, below.

Type of frequency regulation	Incentive for Provision
Primary Frequency Response	Mandatory requirement
Regulation FCAS	Offered service to AEMO
Voluntary frequency response	Rewarded through FPP[8]

Table 5: types of frequency regulation

In 2022, the AEMC made a final rule to introduce Frequency Performance Payments (FPP) which is due to be implemented in 2025. As discussed below, they are used in the proposed design to provide incentives for retailers to make their DR visible.

Pricing errors will lead to greater quantities of frequency regulation being drawn upon through these channels, and so higher costs. Depending upon the source of frequency regulation, these costs may be borne by generators or retailers/customers.

It is plausible that gross pricing errors – and resulting imbalances - could lead to these resources being exhausted, with consequential security impacts such as load shedding. This will be avoided so long as AEMO schedules larger amounts of regulation FCAS to cover these extreme conditions, but this will add further to costs and cause these services to be poorly utilised.

Indirect Impact of Demand Response on AEMO Base Demand Forecasting

DR also has an indirect impact on dispatch efficiency and security, by degrading the accuracy of AEMO's demand forecasting. That is because AEMO's forecasting methods have no model of demand elasticity and will naturally interpret any changes in demand caused by this elasticity as changes in the base demand level.

Demand forecasting methods can be univariate or multivariate. Univariate forecasting involves extrapolating the historical demand time series using statistical methods, whether conventional or machine-learning. Multivariate methods incorporate exogenous factors into the forecasting model, such as ambient temperature. So, for example, a model of the impact of temperature on demand would be defined and its parameters estimated; this model can then be used to remove the temperature impacts from the demand history to create a *weather-corrected* demand series. Univariate methods can then be applied to this weather-corrected series, to accurately forecast weather-corrected demand. Weather forecasts are then applied to the model to add back weather impacts to the weather-corrected demand forecasts, creating demand forecasts that properly reflect forecast weather.

Spot price similarly affects demand through demand response. In a multi-variate analogy, the "price-corrected" demand is simply the base demand, calculated by adding demand response to the metered demand. But, of course, this process requires DR to be visible. It is currently invisible, and so no price-correction is possible, leading to adverse impacts on forecasting accuracy. In short,

---

[8] Strictly speaking, *all* sources of frequency regulation are rewarded through FPP. However, this is less relevant for the first two sources, because these would be required to provide this even in the absence of any FPP incentives.

AEMO's operational forecasting model is missing one of the key factors that drive demand: spot prices.

Figure 5 shows what would happen under a very simple univariate demand forecasting model where demand is assumed to be flat, so the five-minute forecast simply extends the current demand level[9]. Demand is first forecast at D_t at time t based on actual demand in the prior dispatch interval. This sets off the following dynamics:

A. Dispatch sets a spot price, P_t, at the point where the base demand forecast, D_t, intersects the supply curve.
B. Demand response means actual demand for this DI, at the intersection of P_t with the demand curve, is lower than this forecast. This lower actual demand sets the demand forecast for the next dispatch, D_{t+1}.
C. The lower demand now leads to a lower clearing price P_{t+1}.
D. This price fall induces a higher demand than expected, which then carries forward into the next demand forecast D_{t+2}
E. This higher demand forecast now leads to dispatch setting a higher spot price P_{t+2}. And so on.

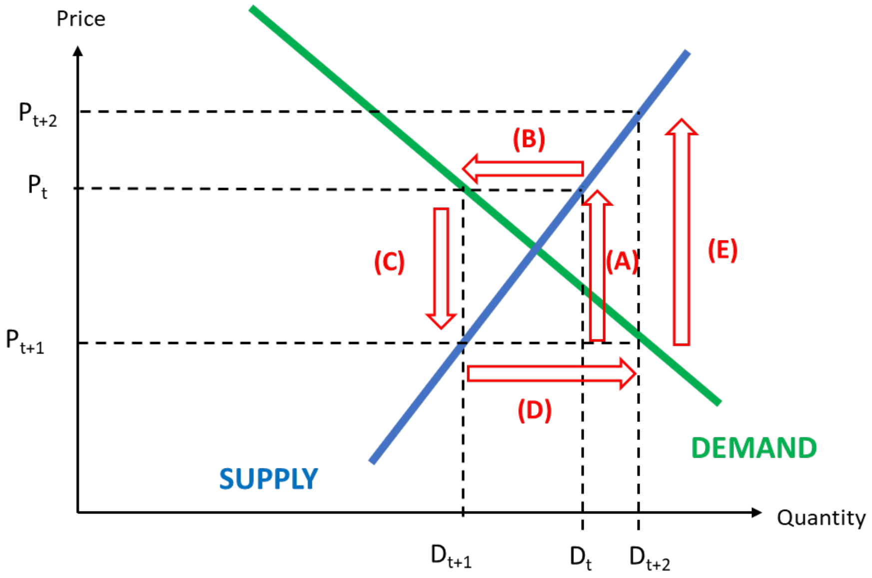

Figure 5: A hog cycle in Dispatch

---

[9] AEMO's forecasting model will be more sophisticated than this, of course, but this assumption helps to simplify the exposition.

This unstable pricing behaviour is known as a hog cycle, named after a situation where supply (of pigs rather than electricity) is a lagged response to price[10]. It is, of course, a worst-case scenario. The oscillation only spirals outwards because demand is more elastic than supply, although this is plausible in electricity when spot prices are very high. Furthermore, AEMO's demand forecasting algorithm will include a damping method to stop such oscillations. Nevertheless, it illustrates the twin impacts of invisible DR on demand forecasting:

- an *initial* effect of the spot price leading to demand response and so a forecasting error;
- a *rebound* effect of this demand response being incorporated into the demand history and then adversely affecting future demand forecasts.

There are corresponding twin benefits of making DR visible and incorporating it into the forecasting and dispatch processes.

Conclusions

AEMO's dispatch process is essentially an auction tasked with discovering the price at which supply and demand are in balance. To support this, AEMO receives information about the supply side through generator bidding. But it has no corresponding knowledge of the demand side, only the historical aggregate regional demands. These have sufficed to date, because there is limited demand response to spot price and so all the necessary information is contained in the demand history. But if DR grows as expected, spot pricing will become increasingly inaccurate, leading to growing imbalances that must be managed through costly frequency regulation.

DR visibility should be encouraged through financial incentives. These don't exist at present, because all demand pays the same price, whether its price response is visible (through scheduled load bids) or not. A fundamental challenge is to create satisfactory incentives for retailers to make their demand response visible, in a form that it can then be incorporated into the dispatch auction.

---

[10] and where the lag is a year rather than 5 minutes

3. Overview of the Design

Overview

The proposed design is based on retailers being financially incentivised to estimate DR and submit these estimates to AEMO so that they can be incorporated into dispatch. Financial incentives are created by adapting and extending existing processes: specifically demand forecasting and frequency performance payments.

This approach is encapsulated in the following design principles:

1. The demand forecasting role is split between AEMO and retailers: AEMO forecasts the base demand (that is, the demand without DR) whilst retailers estimate their DR;
2. Retailers are financially incentivised to provide accurate estimates of DR rather than being mandated to do so;
3. Existing NEM systems and processes are used, adapted or re-purposed to the extent possible: this reduces costs, and enhances transparency and ease of implementation; and
4. Obligations on retailers that choose to estimate DR and submit quasi bids remain light[11] compared to the existing strict compliance obligations on scheduled generation and load.

The selection of these principles is explained for each in turn below.

Division of Forecasting Task

The forecasting of base demand is assigned to AEMO because short-term demand forecasting relies on knowledge of current demand. AEMO has this information through SCADA metering[12]; retailers do not, because customer meter readings are generally not collected until the following day[13]; even if they were available in real-time, aggregation would be a difficult and time-consuming process.

On the other hand, the estimation of demand response is best done by retailers because:

- They are likely to already have relevant and detailed customer information;
- They can be financially incentivised to incur proportionate costs on estimating DR with appropriate accuracy; and
- AEMO estimating DR would likely be contentious: since DR estimates will be a key driver of prices, AEMO could be seen to be effectively deciding spot prices.

The question of retailers also forecasting their base demand is considered further in chapter 8.

Incentivisation

Financial incentives encourage an efficient trade-off between the costs of accurate DR estimation and the benefits to the extent that the incentives reflect the benefits. Some retailers may have little or no DR; others might find it very difficult and expensive to estimate the DR they have. In each case,

---

[11] "scheduled lite" is a broad term to refer to new rules that, whilst being more onerous than existing rules for non-scheduled resources, are less onerous than existing rules for scheduled resources. The proposed design is towards the lighter end of this spectrum: ie closer to non-scheduled than scheduled.

[12] in fact, AEMO only has information on generation and on interconnector flows, but infers regional demand from this

[13] and later for non-smart meters

the cost of DR estimation could exceed the benefit and it would be better – for the retailer and for the market as a whole - if it were not undertaken at all. Similar logic applies to the amount of extra effort and cost involved in improving the accuracy of the estimates.

These incentives also provide for a gradual but timely development of DR estimation as the quantity and materiality of DR grows over time.

Existing systems

Making use of existing systems is obviously desirable, but not always possible. However, the proposed design does allow for existing systems to be retained and/or adapted as follows:

- *NEMDE*: DR estimates can be incorporated into NEMDE by structuring them like *bidirectional unit*[14] (BDU) bids which NEMDE uses currently;
- *Demand forecasting*: Retailer-level demand forecasting (discussed below) can be based on existing AEMO regional demand forecasting models and systems;
- *FPP scheme*: Incentives can be defined and calculated using existing FPP mechanisms, expanded and enhanced accordingly.

This approach should minimise the cost of implementing and operating the proposed design, discussed further in chapter 9.

Lite Compliance Obligations

DR can already be made visible through scheduled load bidding. However, retailers and customers rarely choose this route, due to the compliance costs involved. To be effective, compliance obligations associated with visible DR must be made as "lite" as possible. Compliance should be the minimum necessary to achieve the design objectives of better spot pricing[15]. Obligations proposed for DR are compared with those existing for scheduled load in table 6 below.

	Scheduled Load	**Visible DR in Proposed Design**
Submit bids	Yes	Yes
Provide SCADA metering	Yes	No
Receive dispatch instructions	Yes	No
Follow dispatch instructions	Yes	No
Constrained-on or -off	Potentially	No

Table 6: compliance obligations for scheduled load and visible DR

This principle of minimal compliance will help to encourage DR visibility.

---

[14] a bidirectional unit is a scheduled resource that can both consume and generate electricity, such as a battery

[15] It is acknowledged that other scheduled lite designs might have broader objectives than this, necessitating a somewhat "heavier" approach.

Design Architecture

The proposed design is quite complex and has many moving parts. However, its fundamentals are straightforward, and can be distilled down into four steps, as presented in figure 6, below.

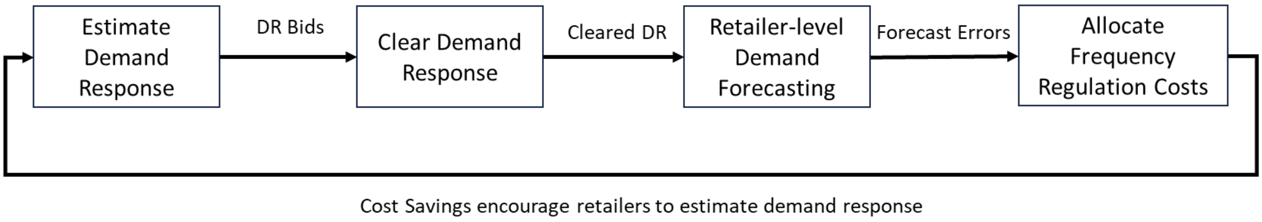

Figure 6: Four Steps are used to create incentives for visibility of demand response

The key steps are as follows:

- *Step 1: Estimate DR*: retailers can estimate their customer DR and submit this information to AEMO via quasi-bids;
- *Step 2: Clear DR*: DR information is fed into NEMDE, as BDU bids are today, spot prices are set and cleared DR amounts determined;
- *Step 3: Forecast Retailer demand*: after real-time, AEMO feeds retailer demand actuals and cleared DR amounts into its forecasting systems to mimic five-minute ahead demand forecasting for each retailer and associated forecasting errors;
- *Step 4: Allocate frequency regulation costs* by adapting and extending the existing FPP scheme to take forecasting errors into consideration.

Where a retailer bids its DR, the cleared DR amounts are then used by AEMO to improve forecasting of retailer demand. This leads to lower forecasting errors, meaning lower costs allocated to the retailer.

Of course, the actual design is rather more complicated than this. A more detailed design architecture is presented in figure 7 below.

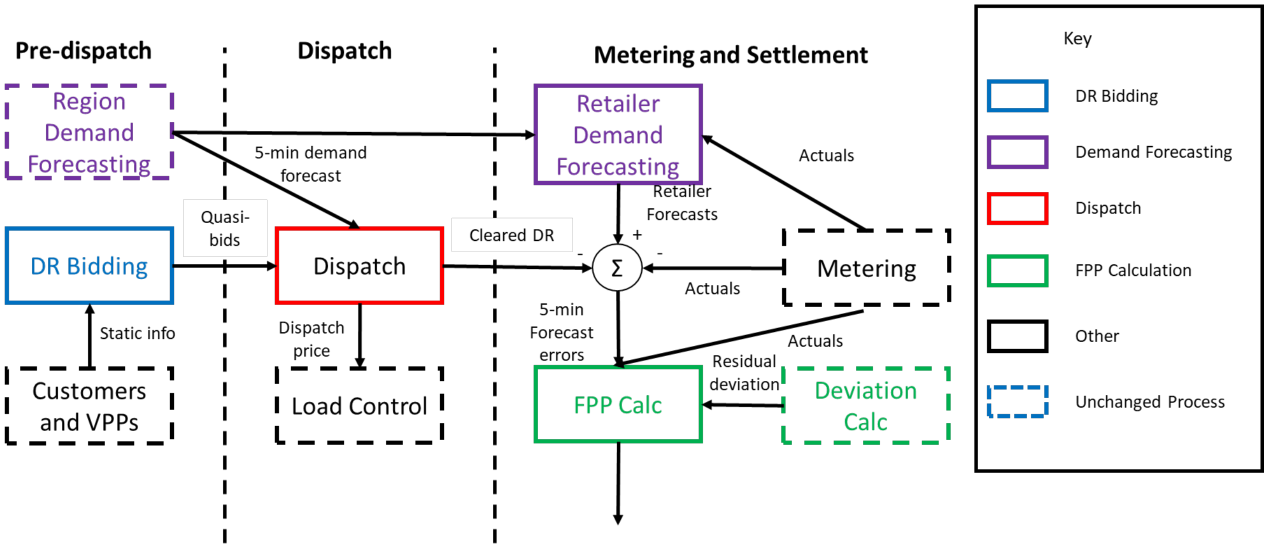

Figure 7: high-level architecture of Proposed design

The main processes are described in turn in the following chapters:

- Chapter 4 describes the estimation and bidding of demand response (blue box);
- Chapter 5 describe the dispatch of demand response (red box);
- Chapter 6 describes the new *retailer-level demand forecasting* process (solid purple box);
- Chapter 7 describes changes and extensions to the FPP process (solid green box).

Conclusions

The proposed design principles lead to a design architecture in which retailers will voluntarily estimate and bid their demand response in order to reduce their demand forecasting errors and so reduce the amount of their contribution to frequency regulation costs.

4. Demand Response Estimation and Quasi-Bidding

Overview

Invisible DR leads to spot prices that cause an imbalance between supply and demand. Invisible DR can also lead to increased demand forecasting errors through the rebound effect described earlier. To make DR visible, the proposed design provides an opportunity and incentive for retailers to estimate their DR and provide these estimates to AEMO in the form of *quasi-bids* that have a similar structure and purpose to bids currently submitted for scheduled resources.

Defining Demand Response

As discussed in the previous section DR is the change in *actual* (ie metered) demand, due to short-term factors such as spot price variability, from the *base* demand that would arise in the absence of these factors. For example, in terms of spot price response, we can think of the base demand being the demand expected if the spot price were to be at some *base price*. Retailers are not required to forecast base demand, although they might decide to do so internally as part of the process of estimating demand response.

The rationale is that AEMO's existing short-term demand forecasting systems should do a good job of forecasting this base demand. However, AEMO will find it hard to model and estimate demand response, given that they will have no knowledge of whether and how customers are responding to spot prices.

In fact, we can turn this around to generalize the meaning of these terms:

- *Base demand* is whatever AEMO is able to forecast, using its current demand forecasting methodologies
- *Demand response* is whatever AEMO is unable to forecast.

So demand response is the missing piece of the puzzle for demand forecasting. Whilst response to spot price is a key (likely the major) factor missing in AEMO's forecasts, response to other factors such as forecast spot price or dynamic network tariffs (such as critical peak tariffs) could also be significant.

Since these definitions are predicated on AEMO's forecasting methodology, it is important that this is transparent and that stakeholders are informed of, and consulted on, any changes to it. For example, if AEMO makes a change to incorporate responses to dynamic network tariffs, retailers will need to know about this so that they no longer incorporate this response into the demand response estimates: to continue to do so would lead to double counting of the response.

Note that "demand response" is a noun, whilst "price-responsive" is an adjective. It is the price-responsiveness of non-scheduled resources that gives rise to demand response.

Estimating Demand Response

The design does not specify how the retailer should go about estimating the demand response of its customers. The process selected will depend in large part on the retailing model, or models, that the retailer employs to encourage demand response from its customers. Some possible models are discussed in appendix A.

Conceptually, demand response is a MW quantity which is a function of price. Since spot prices can range between -$1,000 and +$16,600, the function needs to be defined over this range. The DR function is likely to be constant over some price ranges; it might also be discontinuous at some price points: eg if a retailer arranges to switch off all of its customers' EV chargers when prices exceed $1000, the function will jump between $999 and $1001.

Demand response is referenced to a base demand level, which is the demand at a particular base price selected by the retailer. If spot prices turn out higher than this base price, demand response is positive: ie actual demand is expected to be lower than base demand. If spot prices are below the base price, demand response can be negative: ie actual demand is expected to be higher than the base demand. A possible demand response curve – as a function of spot price – is shown in figure 8, below[16].

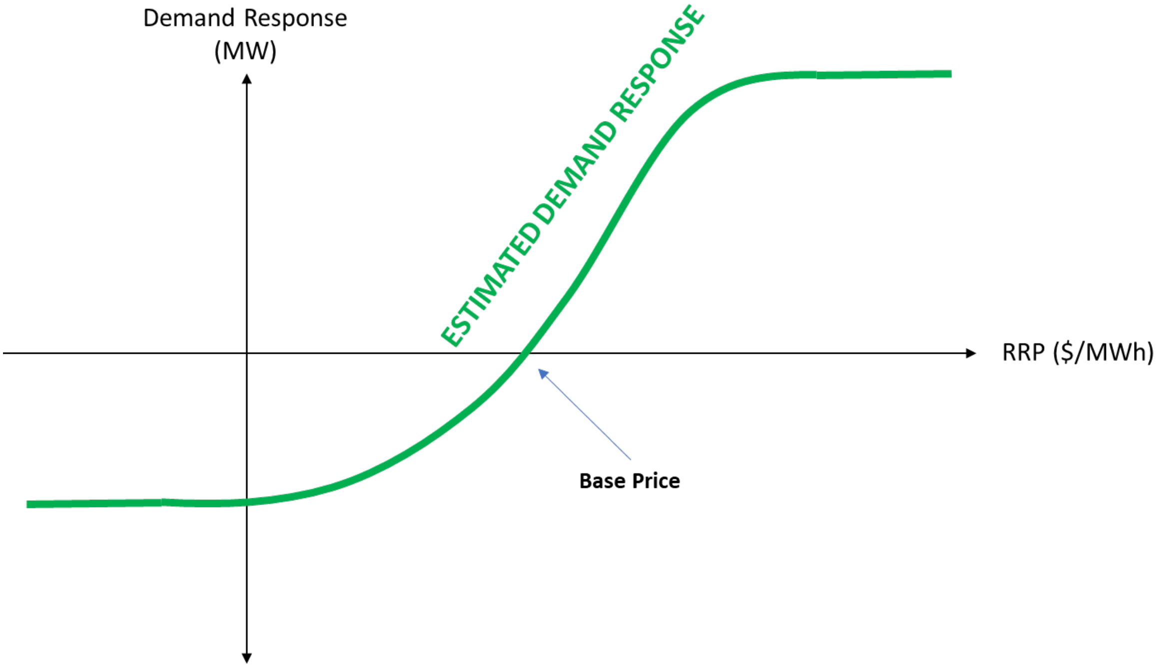

Figure 8: an estimated demand response function

This begs the question as to how the base price is selected. Whilst this could potentially be specified in the Rules or by AEMO, this is probably unnecessary. So long as a retailer is consistent in its choice of base price – by choosing a base price and then sticking to it – the associated base demand will follow a similar pattern to the demand of a conventional retailer (ie where all customers are on conventional fixed tariffs and there is no demand response) and so be able to be forecast by AEMO.

Preparing Quasi-bids

Having estimated its DR, a retailer must send these estimates to AEMO so that these can be integrated into dispatch processes, as discussed in the next chapter. Whilst this information could potentially be presented in any form, implementation will be simpler if the structure used is the same as that used currently for bids from scheduled resources. For this reason, we refer to this transmitted information as *quasi-bids:* "bids" because they look similar to existing bids, but "quasi" as a reminder that these retail customer resources remain non-scheduled.

---

[16] note that the graphs in this chapter have price on the x-axis, being the independent variable. Graphs in Chapter 2 have price on the y-axis, which is the convention when presenting supply-demand diagrams

As noted, the DR can be positive or negative, depending upon the price level. AEMO has recently introduced a new bid structure for bi-directional units (BDUs) – scheduled resources that can operate either as a generator or load. Therefore, it has been decided to use this BDU bid structure for retailer quasi-bids too. This is a pragmatic design choice. Many other structures are possible, but these would likely be more expensive to implement, without bringing any significant advantages.

A BDU bid is made up of 20 price bands: 10 *load bands* cover the load portion, and 10 *generation bands* cover the generation portion. Just like a conventional scheduled load/generation bid, each band has an associated offer quantity and offer price. Offer prices must be monotonically increasing in the load bands and also monotonically increasing in the generator bands. It is permissible for these offer prices to overlap (eg some load offer prices can be higher than some generation offer prices) but only where the associated offered quantity is zero. Offer prices for the subset of bands that have non-zero offered quantities are *not* permitted to overlap – to avoid the possible situation of both generation and load bands being dispatched at the same time.

NEMDE interprets a BDU offer as a response of preferred output to RRP as a "staircase" function with:

- The position of each riser in the staircase being the offer price for the relevant band;
- The height of each riser in the staircase being the offered quantity for the relevant band

The retailer's task is to prepare a BDU-style bid that bet approximates to the DR curve it has estimated. The BDU structure means that it can use up to 10 "steps" above the base price and 10 "steps" below the base price. It might therefore select a base price which allows it to make best use of the 20 steps available. Figure 9, below, shows an example of a bid-staircase matched to the DR curve previously presented in figure 8, above.

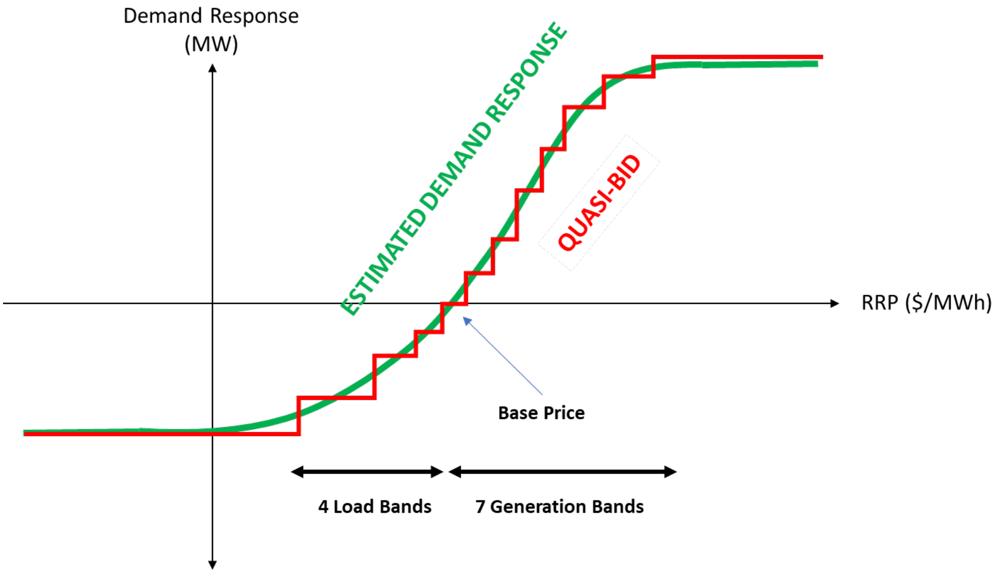

Figure 9: quasi-bid to approximate to the estimated demand response

A numerical example of a BDU-structured quasi-bid is shown in table 7 below. Note that only 4 load bands and 8 generation bands have been utilized. The remaining bands would have zero offered quantity and so are not shown in the table. The bidding retailer has chosen a base price around $100-200/MWh. Below that price, DR will be negative (outturn demand is expected to be higher than base demand); above that price, DR will be positive (outturn demand is expected to be lower than the base demand).

	Offer Price ($/MWh)	Offered Quantity	Cumulative Quantity
L1	5	-100	-190
L2	10	-50	-90
L3	20	-40	-40
L4	100	-20	-20
G1	200	100	100
G2	500	100	200
G3	1000	200	400
G4	2000	500	900
G5	5000	500	1400
G6	10000	300	1700
G7	15000	200	1900
G8	17000	250	2150

Table 7 Example of a quasi-bid of demand response

Under bidding rules, initial bids are submitted day-ahead. Offered quantities can be changed via rebids closer to dispatch, but offer prices cannot. It is proposed that equivalent rules apply to quasi-bids, and the retailer will need to take that restriction into account when choosing their offer prices.

Also, under existing bidding rules, offered quantities must be in whole MW: ie no decimal places are permitted. This may offer insufficient accuracy for very small retailers, whose maximum DR might only be a few MW. It is therefore proposed that decimal places are permitted in quasi-bid quantities.

Quasi-bidding of other Types of Demand Response

The construction of quasi-bids discussed above would apply when a retailer has demand that is responsive to actual spot prices. However, as discussed in appendix A, there can also be demand response which is driven by other factors, such as *forecast* spot prices or dynamic network prices. By our definitions, this is still demand response: ie a customer response to short-term factors not captured in AEMO’s forecasting models. However, because the response is the same irrespective of the actual spot price, the DR curve is flat, and the corresponding quasi-bid would require a single band, with the estimated DR quantity bid at the market floor price (-$1000/MWh). This will be a load band or a generator band in the BDU bid, depending upon whether the DR is negative or positive, respectively.

For example, suppose a retailer calls “demand management” on a day, and expects a response of 10MW from 5pm to 7pm that evening. It will submit the quasi-bid for 10MW, at the market floor price, to cover dispatch intervals over this period. Outside of this period, no quasi-bid is submitted.

Timing of Quasi-Bids and Rebids

Scheduled bids are required to be made at least a day in advance. Offered quantities, but not offer prices, can then be changed, or *rebid*, at any time right up to real-time. NEMDE will run with the latest available re-bids, so any *gate closure* is only a practical limitation of how long it takes for bids to be submitted, received and made available to NEMDE.

Rebids are also subject to rules that prohibit *false or misleading* rebidding which means rebids can only be made to reflect new information that becomes available to the bidder that is brief, verifiable and specific, was not available at the time of the prior bid, and that it intends to honour (subject to new information coming to light).

It is proposed that quasi-bids and rebids would be submitted for the same timescales, and subject to the same rebidding rules. Of course, retailers who choose not to actively participate in the visibility mechanism do not need to submit quasi bids. Participating retailers who expected to have no DR for a particular day would simply submit a bid with zero quantity.

Conclusions

For convenience of implementation, it is proposed that quasi-bids submitted by retailers have the same structure as the bids submitted by scheduled BDUs. The timing of bids and rebids would be similarly aligned.

5. AEMO's Acceptance and Clearing of Quasi-bids

Overview

Quasi-bids submitted by retailers are received by AEMO, where they will ultimately be submitted to NEMDE to be included in the dispatch calculation. However, before this happens, and to prevent erroneous quasi-bids adversely affecting dispatch outcomes:

- The retailer will first need to *qualify* by satisfying AEMO that its quasi-bids have the requisite accuracy and integrity; and
- The quasi-bids will be subject to an automatic *validation* process.

Quasi-bids will be monitored by AER for compliance with rebidding rules. Retailers found to be in breach could be penalized and also disqualified from submitting further bids to NEMDE.

Quasi-bids submitted to NEMDE will receive the equivalent of dispatch targets, referred to as *cleared DR*. There are no conformance obligations associated with these amounts. However, cleared DR is used to price-correct the demand history to prevent historical DR adversely impacting AEMO's demand forecasting.

The process architecture for acceptance and clearing of quasi-bids is shown in figure 10, below.

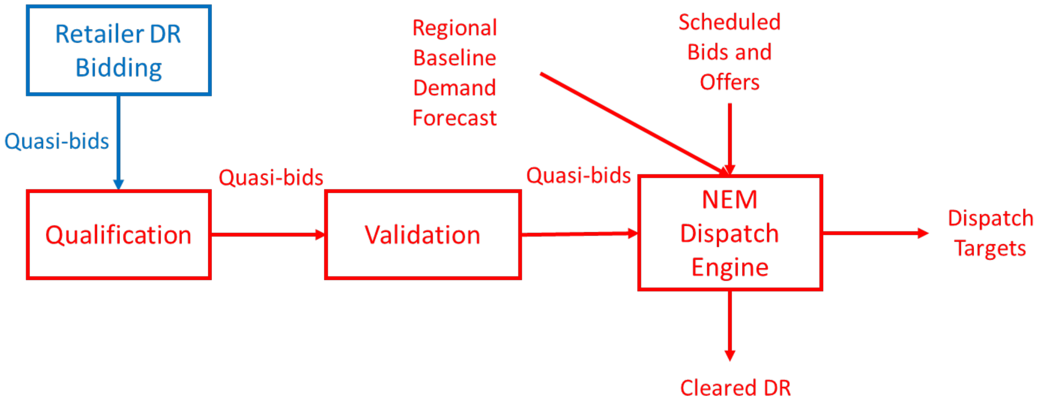

Figure 10: architecture for acceptance and clearing of quasi-bids

Retailer Qualification and Disqualification

Quasi-bids can be submitted by any retailer, but they will only be entered into NEMDE once the retailer has successfully navigated a *qualification period*. During this period, AEMO will monitor the accuracy of the quasi-bids by feeding them into the retailer-level forecasting (RLF) process discussed in the next chapter. The minimum qualifying period will be long enough to encompass a full range of spot price behaviour: indeed, some demand response will be immaterial unless extreme spot prices are encountered. The qualifying period will then continue until sufficient accuracy has been demonstrated.

During the qualifying period, the quasi-bids are not fed into NEMDE and so cannot affect dispatch outcomes. They would initially also not be fed into the FPP calculations, so the retailer would be charged FPP as though its DR remained invisible. However, the retailer could request at any time that the quasi-bids are incorporated into the FPP calculations; it would likely only do this once it was confident that the quasi-bids were accurate and so would lead to lower FPP payments. In any case, once qualified, the quasi-bids would automatically be included.

A similar process could be undertaken for potential later disqualification. If AEMO has concerns that poor quasi-bids from a retailer are adversely affecting dispatch and frequency regulation, it could again use the RLF process to monitor quasi-bidding accuracy over a period, and ultimately disqualify the retailer if accuracy falls below the required hurdle. A retailer would then need to qualify again before its quasi-bids are once more submitted to NEMDE. A disqualified retailer could opt for its quasi-bids to be removed from the FPP calculation.

Bid Validation

Once a retailer has qualified – and unless or until they are disqualified – quasi-bids will be incorporated into dispatch using NEMDE. They will also be used in the calculation of FPP. But before quasi-bids are included in dispatch, they are subject to a validation process.

The RLF process can identify inaccurate bidding, as discussed above. But because this only takes place after real-time, it is too late to be useful to the validation process. Rather, some basic validation rules would be designed and applied. One approach would be for retailers to be required to provide static data – akin to the rated capacity of scheduled generators – that specified the maximum demand response possible from a retailer. This might vary by time of day or day of week etc. Any quasi-bids exceeding this amount could be rejected or capped at the maximum. As with scheduled BDUs, there could be a maximum limit for negative DR as well as positive DR. There might also be an *availability* field in the bid (as with scheduled bids), with the aggregate of the bid quantities not permitted to exceed this availability.

Since quasi-bids are initially submitted day-ahead– with rebidding continuing during the pre-dispatch period – suspect quasi-bids (eg those very different from the same time on the previous day) could potentially be identified and flagged, with the retailer being notified. The retailer could then check the bid and either change it (if an error was found) or reply to AEMO with an explanation of the anomaly. This is necessarily a manual process, and so bid flagging would need to be exceptional rather than routine to make it practical.

Since quasi-bids may include fractions of MW, all offered quantities are rounded to the nearest MW before being entered into NEMDE.

Regulation of Quasi-bidding

By design, DR included in NEMDE will tend to dampen spot price spikes and volatility. A retailer who is short of contract cover might financially prefer these lower spot prices. This might lead it to submit false or misleading quasi-bids: ie bid DR that it knows does not actually exist. Because these quasi-bids are inaccurate, they will lead to larger forecast errors, and so higher FPP charges, for the retailer. However, the retailer might regard this as a price worth paying for reducing spot price exposure.

False and misleading bids from scheduled resources are prohibited currently, with the AER overseeing and enforcing this rule. It is proposed that this rule – adapted as appropriate – also applies to quasi-bids from retailers. It is likely that any quasi bids aiming to material impact spot prices will be highly scrutinised by the AER[17]. Furthermore, the AER could disqualify retailers from submitting further bids to NEMDE, meaning that they could no longer influence spot price outcomes. Such retailers could potentially be permitted to requalify, as discussed above.

---

[17] under NER cl.3.8.22A

NEMDE and Dispatch

With the quasi-bids now submitted to NEMDE just like any other BDU bid, they will be reflected in the dispatch and spot price. The NEMDE output will now include a "dispatch target" for each quasi-bid, which represents the amount of DR that is expected from the relevant retailer at the actual spot price. But retailers are not required to meet to this target in the same way as other market scheduled resources are, because:

- There is no real-time metering of retailer demand, and so nothing for AEMO to use to monitor conformance;
- Even if real-time metering were available, the demand response represents a *difference* from base demand, which is not defined or known in real-time;
- For reasons discussed above, retailers will generally have insufficient knowledge or control of demand response to follow a dispatch target.

However, this "dispatch target" is an important variable for the model, and will be referred to here as the *cleared demand response*. It plays a key role in adjusting and improving demand forecasts, as discussed next.

AEMO Base Demand Forecasting

As discussed in chapter 2, the presence of DR in demand actuals will likely cause rebound effects that degrade demand forecasting accuracy and so induce pricing errors. This can be avoided by correcting the demand history for DR, as shown in figure 11.

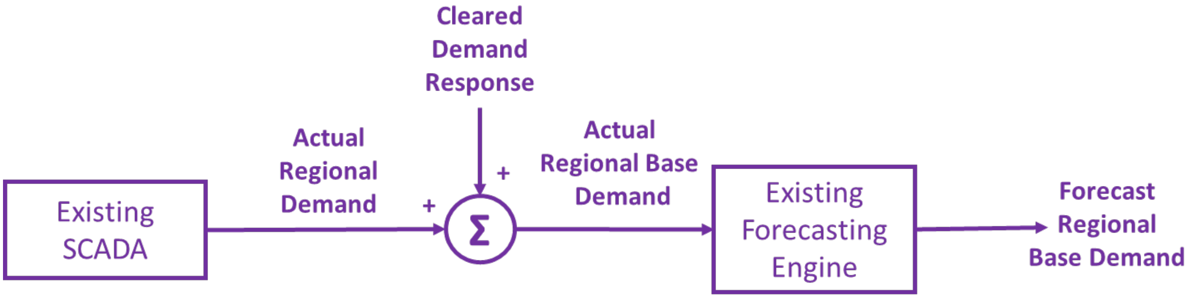

Figure 11: detailed architecture of regional base demand forecasting

Immediately after dispatch, when spot prices have been determined, AEMO will aggregate the cleared DR amounts across each region. This aggregate amount is then added to the actual (metered) demand to determine the (price-corrected) base demand.

The calculated base demand is predicated on the base price used in the quasi-bids. So long as the chosen base price is constant or stable, this will calculate a stable, price-insensitive, base demand series which can be fed into the demand forecasting algorithm that AEMO uses today, to generate reliable and price-insensitive base demand forecasts. This means that the "rebound" effects discussed earlier are neutralised.

For example, consider a situation where, with the demand otherwise stable and flat, the onset of high spot prices[18] triggers substantial demand response. This continues for several DIs before spot prices subside, the demand response disappears, and the demand returns to its previous level. Suppose for simplicity the stable base demand level is 10GW and the demand response is 1GW.

---

[18] perhaps triggered by a major generation outage

Under the current design, AEMO's five-minute forecasts will accurately predict the 10GW of flat demand, but cannot anticipate the DR response and so will predict 10GW for the DI when the DR is first triggered; the actual demand falls to 9GW, and so there is a 1GW forecasting error. Since the actual demands now continue at 9GW for several DIs, the forecasting engine will gradually adjust, and so eventually just predict 9GW. This will be accurate for as long as the DR lasts, but when it ends, it will continue to predict 9GW even as the actual demand returns to 10GW, leading to a *minus* 1GW forecasting error. This is an example of the rebound effect. Again, the forecasting engine will gradually adjust back to the new 10GW level and forecasting accuracy is eventually restored.

In the proposed design, AEMO predicts the *base* demand, of 10GW. When DR commences, the metered demand again falls to 9GW. But the cleared DR is now 1GW and this is added back to the metered demand to give an actual base demand of 10GW. It is this actual base demand which is fed into the forecasting engine. So the forecasting engine will simply see a continuation of the 10GW base demand; the DR effect is removed from the base demand history.

Thus, AEMO continues to accurately predict 10GW of base demand. When the DR ends, the metered demand returns to 10GW but at the same time the cleared DR falls to zero, so the actual base demand simply continues at 10GW and base demand forecasts continue at 10GW. There is no rebound effect.

Pre-Dispatch

Because the quasi-bids are submitted at least day-ahead, like bids from scheduled resources, they can also be incorporated into the pre-dispatch (PD) process. The modifications are similar to those for the dispatch process:

- The BDU quasi-bids from retailers are incorporated into the PD runs, in just the same way as existing BDU bids;
- The historical demand series feeding into the PD demand forecasting systems will be the historical base demand, not the metered demand[19].

Incorporating quasi-bids into the PD process should lead to improve forecasting of spot prices and so lead to more efficient self-commitment and self-scheduling of scheduled resources. It could also help AEMO with decisions around intervention: eg RERT activation.

Conclusions

Because the quasi-bids take the same structure as existing BDU bids, there is no change required to NEMDE or to the dispatch process generally. However, some new qualifying and validation processes would be introduced prior to dispatch.

So long as the spot price is set at the right level, demand will naturally align with supply, so there is then no need for compliance mechanisms to ensure that retailers respond to price as promised. That, in turn, requires DR information provided by retailers to be accurate. This is ensured using financial incentives, through which accurate information is rewarded and inaccurate or missing information penalised. The next two chapters describe how these financial incentives are calculated.

---

[19] Recall that the base demand is calculated by subtracting the cleared DR (from the dispatch process) from the metered demand

6. Retailer-level Demand Forecasting

Overview

Retailer-level demand forecasting (RLF) is a new process introduced in the proposed design:

- To generate five-minute-ahead forecasts of base demand for each retailer;
- Used only for settlement, not operationally;
- Undertaken by AEMO, *not* retailers;
- Carried out after real-time, once retailer actual demands are available[20];
- Nevertheless, designed and specified to *mimic* a real-time forecasting process;
- Making use of AEMO's existing regional demand forecasting models; and
- Using as input the historical, price-corrected, retailer base demand actuals.

The process architecture is shown in figure 12.

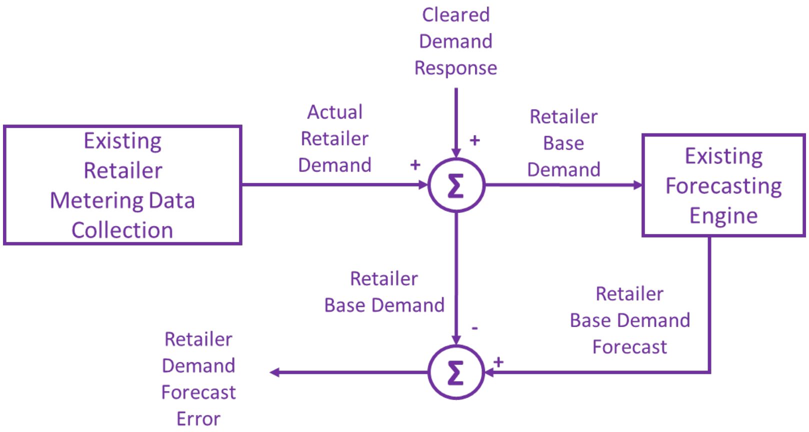

Figure 12: detailed Architecture of Retailer-level Demand Forecasting

Some of these fundamental aspects may be counterintuitive. In other scheduled lite design considerations and proposals, it is generally *retailers* that are responsible for forecasting their own demand. And forecasting, by its nature, is usually carried out ahead of real-time. What is the point of forecasting after real-time, when you already know the answer? And can AEMO use models designed around regional demand to forecast retailer demand, which could have different characteristics? These issues and questions are considered further below.

---

[20] there currently exists an extended process for determining retailer demand for settlement, using a combination of smart meter readings and profiling of advance meter reads. This occurs in several stages, with amounts being updated or corrected as new meter readings come in. This raises the issue of when this retailer-level demand forecasting should be done and whether it should be repeated and updated. This issue is discussed further in Chapter 7.

Comparison between Region-level and Retailer-level Forecasting

AEMO's continuing of operational forecasting at the regional level for dispatch and PD is discussed in the previous chapter. There are similarities, but also key differences between the existing region-level forecasting and the new retailer-level forecasting proposed for the design. These are summarised in table 8 below.

Feature	**Regional Forecasting**	**Retailer-level Forecasting**
Purpose	Used in Dispatch	Assess accuracy of quasi-bids
Who does it?	AEMO	AEMO
When	Real-time	After real-time
Forecasting Horizon	5-minutes ahead	5-minutes ahead
Input data	Generation SCADA	Customer meter readings, aggregated by retailer
Forecasting Methodology	As AEMO decides	Same as regional, or as similar as practical
Correction for DR	Add cleared DR to metered quantities	Add cleared DR to metered quantities

Table 8: Regional vs retailer-level forecasting

The key differences are described and explained in turn below.

Purpose is Settlements, not Operations

Unlike regional demand forecasting, retailer-level forecasting cannot be undertaken in real-time, because the information needed for the forecasting – the retailers' actual demands – is not available until customer meters have been read[21]. Therefore, there are no real-time retailer-level forecasts that can be used operationally.

Rather, retailer-level forecasts are used only for settlements. The logic is that accurate DR estimation and quasi-bidding by a retailer will generally lead to more accurate demand forecasts for that retailer. This effect has been discussed and explained in the context of regional demand forecasting, but it is equally true at the retailer level. Therefore, demand forecasting accuracy is used as a proxy for DR visibility: an indicator of whether or not DR is accurately represented in quasi-bids.

AEMO prepares Retailer-level forecasts after Real Time

Since the forecasts are used in determining settlement payments, it is important that they are done by AEMO to ensure objectivity and auditability.

What does it mean to forecast ex-post? Of course, it is not really forecasting in the traditional sense. Rather, it means mimicking a real-time forecasting process.

For example, at or around 12pm, say, AEMO makes an operational forecast for the regional demand at 12:05pm. It does this using actual regional demand information for the historical period up to and including 12pm. Of course, later demand information is not available at the time of the forecast.

---

[21] Recall that dispatch (ie regional) demand forecasting is based on generation data which is provided in real-time through SCADA. There is no equivalent system for customer data.

Forecasting of *retailer* demand at 12:05pm for that same day would be carried out some time later, in the following days or weeks. But it can similarly be done using actual retailer demand only up to and including 12pm on that day. By that time, the actual for 12:05pm – and subsequent DIs - would be available, but it must *not* be provided to the forecasting algorithm. Using information that wouldn't have been available if forecasting in real-time is referred to here as *cheating*, and must be avoided.

That process determines a demand forecast for one retailer for one DI. The process is then repeated for each retailer and for each DI in the settlement period.

Forecasting Methodology and Models

As discussed, the RLF must be carried out ex-post, one DI at a time. This could potentially be done in a *batch*, with an entire day's worth of 5-minute forecast produced in one go, say, using the repeating process described above. This is a very different process to the once-every-DI forecasting that occurs operationally. This might mean that it is difficult or time-consuming to use the operational software to run this process but, ideally it should be possible to embed the same core algorithm within this batch process. A second-best – but still acceptable – solution would be for AEMO to develop a new algorithm, and associated systems, for the retailer-level forecasting. As far as reasonably possible, this new algorithm should be functionally similar to the regional forecasting algorithm.

Will existing forecasting models be able to be repurposed for this new process? The regional demand model will have been designed to estimate and reflect typical characteristics of regional demand – such as daily and weekly cycles, weather sensitivity, short-term randomness and volatility and so on. So long as retailer demands exhibit similar characteristics, and there is no obvious reason why they wouldn't, the same models should be well suited to retailer-level forecasting. Of course, the prominence of particular characteristics might be different: for example, a retailer to industrial customers will likely see a more stable demand with less daily or weekly variation than seen in the regional demand. But AEMO's forecasting method will include algorithms to estimate the corresponding model parameters that describe these characteristics, leading to bespoke parameter values for each retailer.

Challenges could arise where information used to develop retailer level forecasts is only available at a regional rather than the retailer level. One possible problematic variable is rooftop PV output, discussed further in chapter 8.

Correction for DR

In chapter 5, above, a process was described for removing the impact of DR on the regional demand history by adding back the cleared DR to the metered demand, to recover the base demand. It is the base demand series that is then fed into AEMO's regional demand forecasting process, meaning that it is forecasts of *base* demand that are fed into NEMDE.

A similar process is applied at the retailer level, with each retailer's cleared DR[22] added back to its metered demand (ie the aggregate of its customers' metered demands) to recover the underlying base demand. It is, similarly, the base demand series that is fed into the RLF forecasting process. Forecasting errors are then determined by comparing forecast base demand with actual base demand. This process is illustrated in figure 12, above.

---

[22] note that this cleared DR comes from NEMDE outputs and so will be in whole MW. For small retailers who have submitted quasi-bids in fractions of a MW, there would need to be a process to recalculate the DR directly from the original quasi-bid, to get the necessary precision in the cleared DR amount

If the quasi-bids are accurate, this should lead to better estimates of base demand and so better forecasting performance, given that AEMO's forecasting models are designed to forecast base demand. This better performance then leads to lower FPP charges, as described in chapter 7. So this creates the incentive for retailers to submit quasi-bids. In their absence, the RLF process would simply be based on actual demands.

But how would we even know whether there is such an improvement in forecasting performance? This assessment could potentially be done by producing *two* sets of retailer-level forecasts:

- Forecasting from historical *base* demand: ie metered demand corrected by the cleared DR arising from quasi-bids
- Forecasting from historical *metered* demand: ie ignoring quasi-bids and not attempting to correct for DR.

If quasi-bids are helpful, the former forecasts should be more accurate than the latter. This comparison will be the basis for assessing during the qualification process whether a retailer's quasi-bids are reasonably accurate. The hurdle for qualification would be that forecasting performance is no *worse* when quasi-bids are incorporated.

Apart from this, only one set of forecasts (those with any quasi-bids included) would be calculated in the RLF process.

Systematic and Non-systematic Forecasting Errors

Broadly speaking the sum of the individual retailer forecast errors in a DI will equal the regional forecast error. Forecast errors may, of course, be positive or negative, depending on whether the forecast is higher, or lower, than the actual, respectively. A positive regional forecasting error, say, would imply that *most* retailer forecast errors are also positive; but not necessarily all: some retailers might have negative forecast errors.

More generally, correlations over time between a retailer's forecasting errors and regional forecasting errors will vary, reflecting the extent to which a retailer's demand tracks, or randomly departs from, the regional demand. Statistically, retailer forecast errors can be divided into two components:

- A *systematic error* which exactly follows the regional forecast error, in proportionate terms;
- A *non-systematic error*, which is uncorrelated with the regional forecast error.

For statistical reasons, systematic errors will generally be proportionate to the retailer size, but non-systematic errors will be higher, in proportionate terms, for smaller retailers[23].

It will be seen in chapter 7, below, that it is only systematic forecast errors that attract higher settlement charges. In the absence of invisible DR, these systematic errors will simply reflect retailer size, meaning that settlement outcomes will be similar to today's design[24]. Therefore, small retailers will not be unfairly penalised as a result of having proportionately higher unsystematic forecast errors. Such errors will be settlement neutral.

---

[23] the laws of probability say that the unsystematic error should rise with the square root of retailer size, which means, in proportionate terms, that it *falls* with the square root of size. So a retailer 4 times larger will have twice the absolute error, and half the proportionate error, on average

[24] where charges are allocated in proportion to retailer size, as discussed in chapter 7

Demand response, however, is *always* systematic: all retailers with DR will respond in the *same* direction, decreasing demand if spot prices are high or reducing demand if they are low. This means that forecast errors caused by invisible DR will also be systematic, growing in proportion to the quantity of invisible DR. There is no diversification benefit for a large retailer with large amount of DR, because DR behaviour is not diverse, but rather *orchestrated* by price signals.

In summary, the design is neutral to retailer size, neither favouring nor hindering larger retailers[25].

Conclusions

The test of the accuracy of the demand response information provided by retailers is whether this leads to lower demand forecasting errors. In the proposed design, demand forecasting is split between AEMO and retailers, and these two pieces must be put back together before retailer demands can be forecast and errors determined. Retailer demand information becomes available only after real-time, and AEMO's existing forecasting methods are then used to undertake this analysis. Retailers with invisible DR will be identifiable by the higher demand forecasting errors which will result.

Small retailers, because of less customer diversity, will naturally have proportionately high demand forecast errors, and it is vital that they are not penalised as a result. The proposed design ensures that non-systematic forecast errors, of the type arising for small retailers, are not financially penalised; only the systematic forecast errors symptomatic of invisible DR.

---

[25] of course, large retailers will likely have more resources to prepare and submit DR bids. But such economies of scale are, unfortunately, a fact of life in retailer-land

7. Calculating Incentives

Overview

As discussed, invisible DR gives rise to spot pricing errors which in turn lead to imbalances between supply and demand. This imbalance must be corrected through frequency regulation provided by generators[26]. Each retailer should bear a share of this cost, reflecting the amount of their invisible DR that is causing this cost. Retailer-level forecasting errors are calculated, which indicate the amount of invisible DR that a retailer has.

However, there are some complications which first need to be addressed. Firstly, as discussed above, retailer-level forecasting errors might be systematic (correlated with the overall regional demand forecasting error) or unsystematic (uncorrelated). Only the former, which is caused by invisible DR, adds to frequency regulation costs, so these two error types need to be distinguished in the settlement rules.

Secondly, even in the absence of any invisible DR or 5-minute forecasting errors, supply-demand imbalances can and will occur within a DI, simply due to random load variations, requiring frequency regulation to manage. Allocating these underlying costs to retailers with 5-minute forecast errors would be excessive and unfair.

Similar issues arise for generators too and are already addressed by existing systems: that is, the causer-pays algorithm currently used to allocate regulation FCAS costs, and the frequency performance payments (FPP) scheme[27] due to be implemented in 2025. Whilst there are key differences between generators and retailers, the FPP concepts and algebra can be modified and repurposed for application to retailers.

FPP concepts are discussed in the next section. The adaptation of those concepts to retailer cost-allocation is then developed in the remainder of this chapter.

Current FPP Design

Overview

The architecture of the current FPP scheme is summarised in figure 13, below.[28] The main processes are described in the following sections below.

---

[26] note storage and even load can also provide frequency regulation

[27] The frequency performance payments introduced in AEMC's Primary Frequency Response Incentive Arrangements final rule commence from 8 June 2025.

[28] Note that this section refers to "generators" and "demand" which, in the terminology used in this paper, means scheduled resources and non-scheduled resources, respectively. The FPP Rule actually uses a slightly different dichotomy: resources with 4-second metering versus those without 4-second metering. It is understood that all scheduled resources will have 4-second metering, and that the vast majority of non-scheduled resources will *not* have such metering, so this distinction is not anticipated to materially impact on the proposed design.

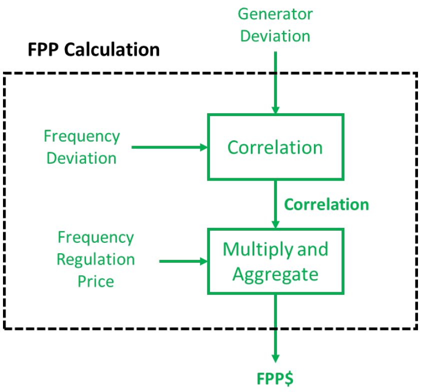

Figure 13: Architecture of the Current FPP Scheme

Generator Deviations

Generators are scheduled[29] rather than forecast; given dispatch targets by NEMDE. Differences, or *deviations*, between these dispatch targets and actual generation will inevitably arise. They are analogous to the forecast errors arising on the demand side.

Dispatch targets are only specified each five-minutes (5M), for the end of each DI. However, to reflect the dynamics of frequency and frequency regulation, the FPP operates at the 4-second (4S) level, meaning dispatch targets and output actuals are required at this granularity. 4S dispatch targets are defined using a 'reference trajectory" which is simply a linear ramp between consecutive dispatch targets. This reflects the dispatch conformance requirement that generators must endeavour to ramp between dispatch targets.

All generators are required to have 4S metering, allowing for active power deviations – the difference between reference and actual – to be calculated every 4 seconds, as shown in figure 14 below.

Figure 14: Calculating Generator Deviations (taken from AEMO's FPP Procedure)

---

[29] Recall that, for simplicity, the term "generators" is used in this report to refer to scheduled and semi-scheduled generators only. Of course, there are also non-scheduled generators in the NEM, which in this report are included under "demand" banner.

Correlations with Frequency

To distinguish between systematic and non-systematic deviations - that is to say between deviations that add to frequency regulation costs and those that don't - the deviations are correlated with a measure of frequency deviation[30].

System frequency reflects supply-demand imbalances: that is to say, aggregate deviations. Therefore, systematic deviations can be distinguished from non-systematic deviations by correlating with frequency deviation. Correlations can be interpreted as follows[31]:

- A *positive correlation* means that a generator is typically over-generating when frequency deviation is positive, or under-generating when frequency deviation is negative, in both cases worsening the frequency deviation and so adding to the costs of frequency regulation.
- For a *negative correlation*, the opposite is the case, and the generator is helping to reduce the costs of frequency regulation.
- A *zero correlation* means the generator is neither mitigating nor exacerbating frequency regulation costs.

A negative correlation is unlikely to happen by chance. Random generator deviations will generally be positively correlated with frequency, because the deviation itself – however small – will impact frequency somewhat. Rather, a negative correlation implies that a generator is *deliberately* providing frequency regulation.

These correlations are therefore used to determine FPP amounts (*FPP$*):

- Generators creating the need for frequency regulation, and so having positive correlation make payments *to* AEMO;
- Generators providing frequency regulation, and so having negative correlation, receive payments *from* AEMO.

Demand Deviation

Customer load does not generally have 4S metering, so it is not possible to calculate deviations directly. However, deviation can be inferred based on the fact that deviations across all participants must always sum to zero. This arises from two simple identities equating supply and demand. Firstly, at the end of each dispatch interval:

dispatched supply = forecast demand

---

[30] frequency deviation is the difference between actual frequency and the nominal 50Hz frequency. The FPP algebra then smooths this using an exponentially-weighted moving average

[31] note that the actual FPP procedure reverses the sign of the smoothed the frequency deviation to come up with a frequency measure. Therefore, in the FPP algebra, *negative* correlation between deviation and frequency measure *adds* to frequency regulation costs. This report will use the usual sign convention for frequency deviation, meaning that a positive or negative correlation exacerbates or mitigates the need for frequency regulation, respectively. This seems more intuitive, but may unfortunately confuse those familiar with the FPP algebra. Apologies.

The FPP creates a reference trajectory for demand in the same way as for generators: ie as a ramp between consecutive 5M points. This alignment with generation dispatch means that, at every point in time:

reference supply = reference demand

Also, the physics of the power system require that, at every point in time:

actual supply = actual demand

The difference between these two equations gives:

actual supply – reference supply = actual demand – reference demand

The LHS is simply the aggregate of all generator deviations. We now define demand deviation as:

demand deviation = reference demand – actual demand

Note that this sign convention is the *opposite* to that for generator deviation, but aligns with the sign convention for demand forecasting error: ie forecast > actual implies a positive forecast error and a positive deviation.

Therefore we get:

demand deviation = -1 x aggregate of generator deviations

Making use of this identity, the demand deviation is simply defined as the negative of the aggregate of the generator deviations, implying that the total of all deviations – demand plus supply – is always zero. This demand deviation is referred to in the FPP rules and procedures as the *residual deviation*[32], but this terminology is not used in this report, to avoid confusion, because the proposed design itself gives rise to more residuals.

Like with generator deviations, the correlation between the demand deviation and frequency deviation is calculated. A positive correlation implies that:

- the deviation is positive (actual below forecast) when frequency deviation is also positive, exacerbating the frequency deviation; or
- the deviation is negative (actual above forecast) when frequency deviation is also negative, again exacerbating the frequency deviation.

So, as for generators, a positive correlation adds to the need for frequency regulation. The demand-side does not provide frequency regulation[33]. so the correlation is unlikely to be negative.

Deviations always sum to zero, meaning that correlations must too, and so the aggregate of negative deviations must equal the aggregate of positive deviations. Put another way, the supply of frequency regulation must equal the demand for it.

---

[32] because the deviation arises from all generation and load *without* 4S metering; whilst this is mostly load, it will include some generation

[33] customers could theoretically deliberately provide frequency regulation, but since they are not required to do so and would not be paid for doing so, this is unlikely. Some rotating machinery has a natural response of load to frequency which does provide some inherent frequency regulation, but this is likely to be outweighed by the effect of other load.

Settlement

FPP settlement algebra is complex, but it essentially works by calculating a price in each DI to apply to the calculated correlation, so:

FPP\$ = FPP price x correlation

Where:

FPP\$ is the FPP settlement amount.

The FPP price is usually positive, so a positive or negative correlation leads to a positive or negative FPP\$, respectively. This means that participants with a positive correlation make payments *to* AEMO; those with negative correlation receive payments *from* AEMO, as a reward for providing frequency regulation.

The FPP\$ amount payable by the demand side is based on the correlation of the demand deviation with frequency deviation:

Demand FPP\$ = FPP price x correlation of demand deviation with frequency

This amount must be recovered, in aggregate, from retailers, and so an allocation method is required. In the absence of 4S metering of customers, it is not possible to ascertain how much each retailer contributed to the demand deviation, so the demand FPP\$ is simply shared between retailers in proportion to their actual demand in the DI. The FPP architecture for retailers is summarised in figure 15, below.

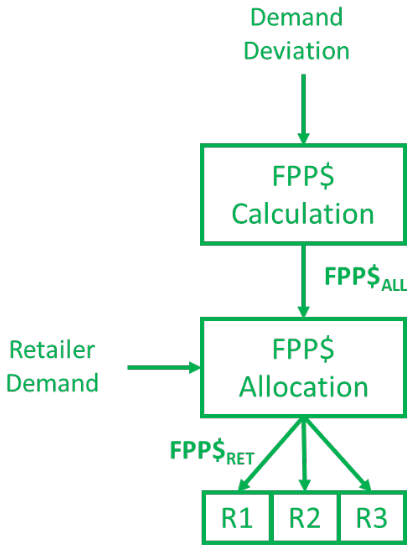

Figure 15: Current Architecture for calculation of retailer FPP Amounts

A simple example of this allocation process is presented in table 9, below. For a DI, the amount FPP\$$_{ALL}$ attributable to the demand deviation is calculated to be \$2000. This is shared between the three retailers responsible for this demand in proportion to their demand in this DI.

Element	Retailer			Total
	A	B	C	
Demand (MW)	1500	200	300	2000
Demand Share (%)	75%	10%	15%	100%
FPP$ ($)	*1500*	*200*	*300*	*2000*

Table 9: Example of Current FPP$

Future FPP Design

Overview

The FPP design is based on the premise that systematic deviations create the need for frequency regulation. On the demand-side, deviations are just demand forecasting errors. The previous chapter showed how five-minute forecasting errors can be inferred for each retailer. The final step is to integrate these "five-minute deviations" with the four-second algebra used in the FPP scheme.

The proposed design does this by adding the following steps to the FPP settlement algebra:

- decompose the demand deviation into three separate components, reflecting the respective impacts of visible DR, invisible DR and random variation of base demand;
- calculate the correlations and FPP$ amounts for each of these three components;
- use different allocation metrics for allocating the three FPP$ amounts between retailers, reflecting the respective drivers of these costs.

This revised design is illustrated in figure 16, below.

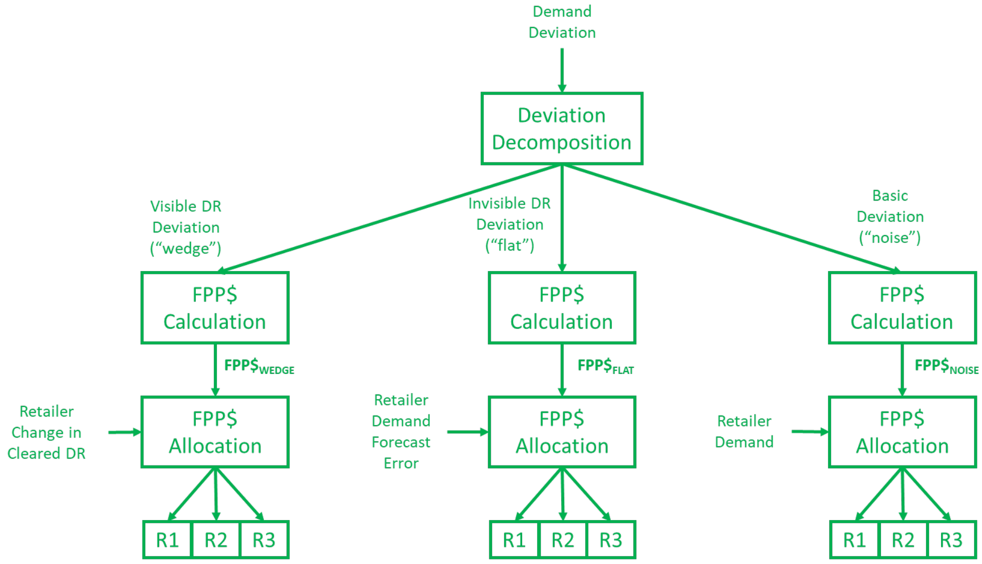

Figure 16: Revised Retail FPP$ Architecture in Proposed design

Decomposing the deviation

The proposed design decomposes the demand deviation into three components, as shown in table 10, below. The components are then described in the following sections.

Component	Caused by	Shape	Size
Wedge	Visible DR	Backward Wedge	Change in cleared DR
Flat	Invisible DR	Rectangle	Estimated from bias
Noise	Base Demand variability	Random Walk	Whatever is left over

Table 10: the three Components of the Demand Deviation

Deviation from Invisible Demand Response

Consider a simple scenario, illustrated in figure 17, below, where base demand is flat and forecastable, but DR is invisible and so unforecastable. AEMO's base demand forecast is flat and, with no DR being bid, there is no cleared DR. So, the 5-minute demand forecasts used to determine the reference trajectory are simply these base demand forecasts, with the reference trajectory being just a line between these two points.

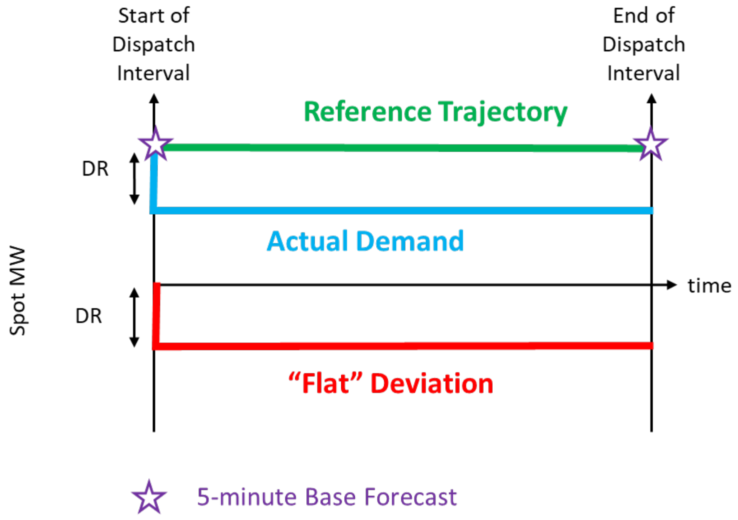

Figure 17: Invisible Demand Response creates a Flat Deviation

Now DR is the response of demand to spot price. Since the same spot price applies throughout the DI, and then abruptly changes at the DI boundary, DR can be expected to behave similarly. So, although the base demand forecast is assumed accurate, the DR creates a step change in actual demand. This in turn creates a step change in the deviation, which is just the difference between the reference trajectory and the actual demand.

In summary, if demand were flat and unforecastable, except for some invisible demand response, we would expect to see a rectangular-shaped deviation: ie a constant deviation over the DI.

Deviation from Visible Demand Response

Similarly, to understand the deviation caused by *visible* DR, consider a different scenario. Base demand is again flat and forecastable, but now the DR is visible rather than invisible: ie accurately estimated by retailers and bid into dispatch. This scenario is shown in figure 18 below.

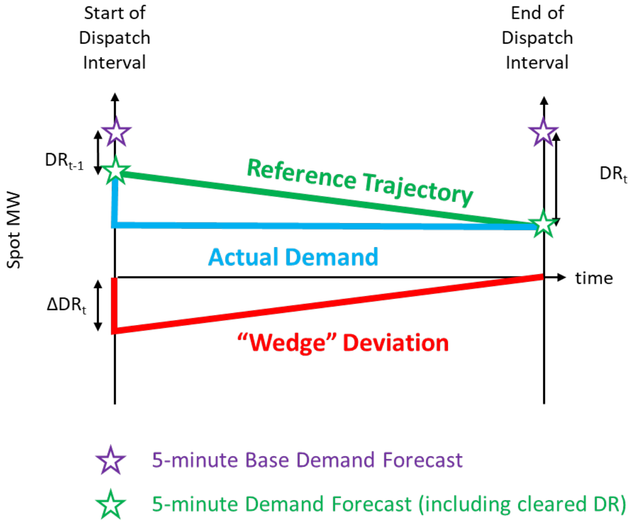

Figure 18: Visible Demand Response creates a Wedge Deviation

Again, AEMO's base demand forecast is flat, but this is now adjusted by the cleared DR. The demand at the start of the DI is adjusted by the DR cleared in the prior DI, DR_{t-1}; the demand at the end of the DI is adjusted by the DR cleared for the current DI, DR_t. Because the spot price changes from the prior DI, there is a corresponding change in cleared DR (*ΔDR*), this leads to different five-minute demand forecasts and the reference trajectory will be a slope joining the two.

The DR is now visible, but will nevertheless have the same response to changing spot prices as the invisible DR: ie a step change in demand at the start of each DI. The deviation is again the difference between reference and actual, but this is now in the shape of a wedge, with its thin end at the end of the DI and the height of its thick end at the start of the DI being ΔDR.

Because the DR is visible, the cleared DR quantities and so ΔDR are known, meaning that this deviation can be calculated precisely for each DI.

Decomposition

In the general case, there is a mix of visible DR, invisible DR and underlying demand variation and volatility. A general method is needed for decomposing the resulting demand deviation into the three components. This method is illustrated in figure 19, below.

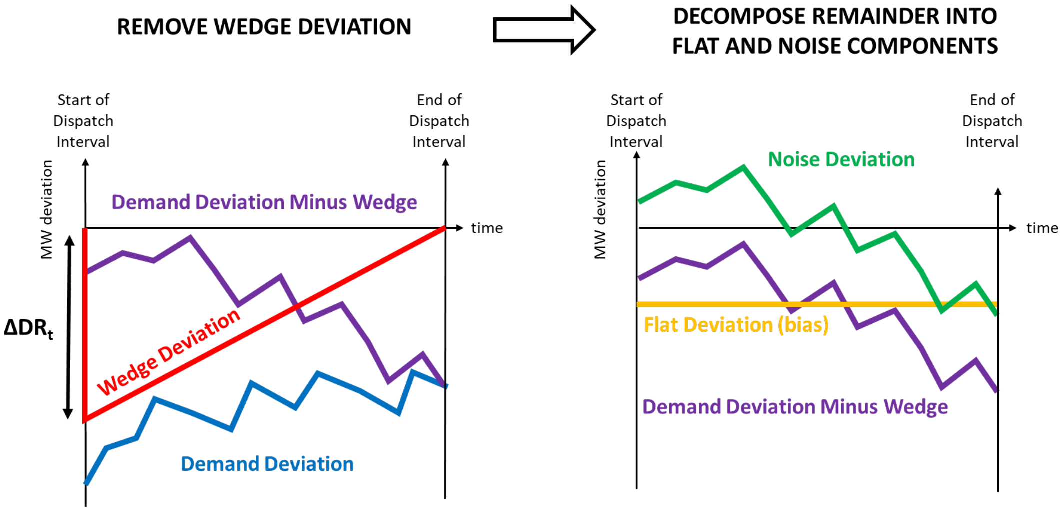

Figure 19: Decomposing the Demand Deviation

The first step is to remove from the demand deviation the wedge component, whose wedge shape and size is known precisely. This leaves a remainder (purple line in the figure above) which must now be decomposed further into the “flat” and “noise” components. The quantity of invisible DR is unknown, but it can be inferred by looking at the average remaining deviation over the DI, or its *bias*, as shown in figure 19 above.

Now this bias cannot be attributed entirely to invisible DR. There would be some bias even if there were no DR at all, just from random demand variations and demand forecasting errors like we see today. But the premise behind the proposed design is that invisible DR will grow substantially and so will ultimately dwarf this underlying randomness[34].

On this basis, the flat deviation is a simple rectangle (a constant deviation over the DI) with its height set at the bias in the remaining deviation. The noise component is then what is left after this flat deviation is removed. As shown in figure 19, this is just a random deviation with zero bias.

---

[34] and, as discussed below, if DR doesn’t grow in this way, this decomposition based on bias will lead to settlement outcomes similar to the current FPP anyway

This decomposition method is illustrated numerically in table 11 below. With 4S metering, there are 75 demand points within each 5-minute dispatch interval. However, for simplicity, only 10 points are used in the example.

Time Interval	Demand deviation	Wedge Component	Remainder	Flat Component	Noise Component
1	-243	-90	-153	-163	10
2	-212	-80	-132	-163	31
3	-210	-70	-140	-163	23
4	-175	-60	-115	-163	48
5	-196	-50	-146	-163	17
6	-221	-40	-181	-163	-18
7	-201	-30	-171	-163	-8
8	-205	-20	-185	-163	-22
9	-234	-10	-224	-163	-61
10	-184	0	-184	-163	-21
Average			*-163*	*-163*	*0*

Table 11: numerical example of deviation decomposition

In the example, the demand deviation is generally negative (ie demand actuals are less than the reference based on five-minute regional demand forecasts) implying that there is significant DR in the DI. Indeed, it is known in this case that ΔDR=100, so the wedge deviation amounts change linearly from -100 to 0 over the DI[35]. This wedge deviation is subtracted from the demand deviation to give a remainder which must then be further decomposed into bias and noise.

The average of the remainder across the DI is -163MW. This is fully attributed to invisible DR, creating a constant “flat” deviation with this amount. This is subtracted from the “remainder” to leave just the noise component which, by definition, has an average of zero.

Settlement

With the demand deviation decomposed, the existing FPP algebra is used to calculate FPP$ amounts for each component. Because the FPP$ algebra is linear[36], being based on correlations with frequency deviation, the sum of these three FPP$ amounts will be mathematically identical to the demand FPP$ calculated in the current design:

Demand FPP$ = wedge FPP$ + flat FPP$ + noise FPP$

All that remains is to allocate these three FPP$ amounts between retailers. A different allocation method is applied to each component, as discussed below.

---

[35] it is 100 at time=0 which is taken to be the end of the prior DI and so not included in the table

[36] a function is said to be linear if the function of the sum equals the sum of the functions. That is f(a+b) = f(a) + f(b)

Visible DR

The visible DR FPP$ arises from the visible DR deviation which, in turn, arises from the ΔDR. Therefore, the wedge FPP$ amount is simply allocated in proportion to the ΔDR of each retailer.

Invisible DR

Chapter 6 described how retailers with invisible DR will have proportionately high *systematic* demand forecasting errors. On the other hand, in the absence of invisible DR, these forecasting errors will be proportionate to retailer size.

Reflecting this theory, the flat FPP$ is allocated in proportion to the retailer demand forecasting error. If there is no invisible DR, this will lead to an allocation in proportion to retailer demand: ie the same as the current design. However, retailers with invisible DR will have disproportionately high forecast errors and so incur relatively high FPP$ charges. This creates the financial incentive for the retailer to make its DR visible through quasi-bidding[37].

Random Variation

Randomness in demand gives rise to the noise component and associated FPP$. This is simply allocated in proportion to retailer demand, mirroring the current FPP design.

Non-systematic Demand Forecasting Errors

This alternative FPP$ algebra provides a fair and efficient allocation in relation to systematic demand forecasting errors. But what about non-systematic ones? As discussed in chapter 6, small retailers will have disproportionately high non-systematic errors, and it would be unfair if this led to them incurring disproportionate FPP$ charges.

By definition, non-systematic forecast errors are as likely, for a particular DI, to have the *opposite* sign to the aggregate forecast error as the same sign. The FPP$ algebra ensures that retailers having the *opposite* sign in a DI receive a *negative* FPP$ in respect of the invisible DR. Therefore, these "swings and roundabouts" will average out over time, and small retailers will not be disadvantaged.

Smaller retailers will have a relatively noisy demand profile at the 4-second level. Of course, this cannot be seen, but nevertheless can create frequency regulation costs. However, only systematic variations create these costs, and these grow in proportion to retailer size. For that reason, the FPP$ reflecting the noise component is allocated in proportion to retailer demand.

---

[37] This will mean it is charged for visible DR instead. However, this is expected to be much lower than its invisible counterpart, for reasons discussed in chapter 8

Numerical Example

A numerical settlement example is shown in table 12 below. This takes the \$2000 FPP$\$_{ALL}$ amount from the previous example shown in table 9 and reallocates it based on the proposed design.

Element	Retailer			Total
	A	B	C	
Features	visible DR	invisible DR	no DR	
Demand (MW)	1500	200	300	2000
Demand Share (%)	75%	10%	15%	100%
Forecast Error (MW)	30	80	-10	100
Error Share (%)	30%	80%	-10%	100%
ΔDR (MW)	20	0	0	20
ΔDR share (%)	100%	0%	0%	100%
Wedge FPP$ ($)	200	0	0	200
Flat FPP$ ($)	450	1200	-150	1500
Noise FPP$ ($)	225	30	45	300
Total FPP$ ($)	*875*	*1230*	*-105*	*2000*
Current FPP$ ($)	*1500*	*200*	*300*	*2000*

Table 12: FPP Settlement Example

As before, the example has three retailers, and the proposed design has revealed these to have different DR characteristics:

- retailer A has substantial DR, which it makes visible by bidding it in;
- retailer B also has substantial DR which it doesn't bid it in and so remains invisible;
- retailer C has no DR of either type.

The FPP decomposition and calculation processes have, for this particular DI, calculated the FPP$ amounts, shown in the green-shaded cells. The demand, forecast error and ΔDR amounts for each retailer have also been calculated and are shown in the orange-shaded cells. Together, these cells contain all the inputs to the process for allocating the FPP$ for the demand deviation between the three retailers.

Recall that, under the current design, the FPP$ for the demand deviation is simply shared between the retailers in proportion to demand (as shown in the bottom row of table 12). So retailer A, being the largest retailer picks up most of the FPP$ charge currently.

In the proposed design, the demand deviation is decomposed and the FPP$ amounts calculated for each component. In the example, the FPP$ for the flat deviation – caused by invisible DR – is largest. This is caused by retailer B, who as a result has the proportionately highest demand forecast error and so pays a higher charge than currently.

Retailer A has made its DR visible, so its forecast error is relatively small and it picks up a smaller change of the flat FPP$. On the other hand, being the only retailer with visible DR, it picks up all of the wedge FPP$ charge. But this is fairly small, because the ΔDR itself is fairly small: perhaps because the spot price has not changed substantially from the prior DI.

Retailer C has no DR, but because it is fairly small it nevertheless has a moderate forecast error. However this error, being random, is non-systematic and, in the example, it has the *opposite* sign to the aggregate forecast error, meaning that it pays a *negative* part of the flat FPP$. As a result, it has a negative charge overall. However, this is purely random and could be positive in other DIs.

The overall outcome is a significant shift in charge allocation to retailer B compared to the current design. This might encourage retailer B to bid its DR in order to reduce its FPP$ charge.

Settlement Timing and Re-runs

As discussed above, the settlement amounts for each retailer associated with the new FPP incentives depend upon the following market outcomes:

1. the demand deviation, its components and the associated FPP$ amounts
2. the retailer's cleared DR – to allocate the wedge FPP$
3. the retailer's demand – to allocate the noise FPP$
4. the retailer's demand forecasting errors – to allocate the flat FPP$

Items 3 and 4 depend directly or indirectly on retailer demands, which are calculated by aggregating the metered demands of their customers. The timing of the processes for reading meters and applying profiles to non-interval meter reads means that retailer demand quantities used in NEM settlement will be provisional initially and then updated iteratively as more meters are read. This raises the question as to whether the FPP settlements process should similarly be re-run and updated when revised retailer demands are available. This could be problematic and resource intensive, particularly if it means re-running the retailer-level forecasting process multiple times.

A key point to consider here is that customers providing demand response will all have interval metering: without this, the value of the DR cannot be captured and so the customer cannot be rewarded. Given this, the amount of DR – whether invisible or visible – will be estimated accurately once the interval meter readings are available. Therefore, there should be no reason for multiple re-runs of the calculation of item 4, above. Whilst the reading of non-interval meters will make a small change to item 3, above, this is easy to recalculate if needed, and the amounts involved are in any case likely to be small[38].

Item 1 is determined from readings of generation SCADA meters and frequency meters; customer meters are not involved. Item 2 is based on cleared DR which is calculated by NEMDE in real-time and never revised. Therefore, no new issues arise around settlement timings for these items.

---

[38] it is unclear whether these re-runs would be done for the existing FPP settlements, which also rely on retailer demand

Conclusions

Frequency deviations arise from systematic deviations from forecasts, creating the need for corresponding frequency regulation to contain frequency within secure limits. The existing FPP scheme identifies these systematic deviations and charges the deviating party accordingly.

On the demand-side, deviations in a DI arise from a base of five-minute demand forecasting errors with the high-frequency noise of random customer activity superimposed. The FPP algebra can be adapted to separately calculate the costs associated with these respective deviations. Since five-minute retailer-level forecast errors are now known, the costs of these can be allocated accordingly. This is the final step in establishing visibility incentives, since DR visibility leads to lower forecasting errors and so lower FPP charges.

8. Design Assessment and Discussion

Overview

The proposed design has been described and discussed over the previous five chapters. This chapter raises and discusses some potential concerns or issues that might be raised by stakeholders. It also discusses areas where the design might be able to be amended or extended, either initially or in the future.

Questions discussed are:

- Does the proposed design only affect retailers?
- Can AEMO's existing regional demand forecasting methods be repurposed?
- Are all responses of demand to price regarded as demand response?
- Can retailers accurately estimate demand response?
- Can retailers forecast their base demand too?
- Should visible DR face a specific FPP$ charge?
- Should visible DR enjoy lower regulation FCAS costs?
- How are retailers with 4S metering managed?
- How does distribution congestion affect the proposed design?
- How does transmission congestion affect the proposed design?

These questions are considered in turn below.

Does the Proposed Design only affect Retailers?

As explained in Chapter 2, for reasons of simplicity this paper uses the term retailer to cover all market participants who have financial responsibility for non-scheduled resources. Retailers are responsible for customer demand which, by definition, is non-scheduled. Whilst we conventionally think of customers simply having load, the picture has become more complicated with customers now commonly having generation in the form of rooftop PV or the discharge of home batteries or even EV batteries using *V2G* technology. At times, this can cause customer demand – and even retailer demand – to become negative: ie because the generation exceeds the load. This is all accommodated by the design. Demand response could as easily be an increase in customer generation in response to higher spot price as a reduction in customer load, but this makes no difference to the design.

Financial responsibility for non-scheduled generation can also be assigned, under the current rules, to a *small generation aggregator*. Since this participant is responsible only for generation, its "demand" will always be negative. Again, this is accommodated by the design, with "demand response" then simply being a response of the generation to changes in spot price.

Can AEMO's existing Regional Demand Forecasting Methods be repurposed?

The proposed design has been developed on the assumption that AEMO's method of regional demand forecasting can be repurposed and applied to ex-post retailer-level demand forecasting. This assumption is based on the forecasting method requiring the same or corresponding inputs and the characteristics of the different demand histories being similar.

There are two potential objections. The first is that the regional and retailer demand series are actually quite different: the regional demands are *spot MW* at each DI boundary, whereas the retailer-level demands are *average MW* over the DI. It is not expected, though that this difference will adversely impact forecasting practicalities or accuracy.

The second potential concern is around rooftop PV, which has become a key component of demand, in the sense that it is netted off customer consumption before the meter and commonly leads to negative demand (ie exports) at the customer level and even the retailer level.

It is understood that, in its regional demand forecasting, AEMO separately forecasts regional rooftop PV output, using static information on rooftop PV installations, together with real-time sampling of rooftop PV output at select installations. Applying this methodology at a retailer-level would likely require this PV information to also be at the retailer level. It is not known whether the information is currently available to AEMO in this form and, if not, whether there would be practical difficulties in obtaining this.

Apart from conceptual concerns, there may also be operational challenges in this repurposing. Regional demand forecasting is an operational tool working within a real-time environment. Retailer-level demand forecasting would operate as part of settlements. At the very least, this repurposing would require a migration of the forecasting applications to a different IT platform, to be operated by a different AEMO department.

Are all Responses of Demand to Price regarded as Demand Response?

As discussed in chapter 3, DR is defined generally to be *any* response of non-scheduled resources to short-term factors not modelled in AEMO's demand forecasting algorithms used for pre-dispatch and dispatch.

This will certainly include response to spot prices, but would also include response to forecast spot prices, as discussed in Appendix A. A grey area would be response to dynamic network tariffs, such as critical peak tariffs. These are notified to customers in advance and they could potentially be notified to AEMO (in the future) for it to incorporate into demand forecasts. This response would then not be regarded as DR, because to do so would lead to double counting, with both AEMO and retailers separately predicting the response. On the other hand, if AEMO did not incorporate this response, it would be left to retailers to include it in their quasi-bids.

In the future, we are likely to see new tariff structures and retail products which further blur the delineation around demand response. This does not matter for the proposed design[39], but might place challenges on the retailer on whether and how to quasi-bid this response into AEMO.

Can Retailers accurately estimate Demand Response?

The proposed design is predicated on retailers being reasonably able to accurately estimate DR in their customer base. If this is impossible or impractical, then the incentives created for them to quasi-bid their DR will be ineffective.

Appendix A describes several possible retailing models through which retailers and their customers could realize and capture the benefits of demand flexibility. These models give the retailer different degrees of visibility and control over the response of the associated non-scheduled resources. Other things being equal, greater visibility and/or control will allow retailers to better estimate demand response and reflect this in their quasi-bids.

Of course, the design does not require that quasi-bids are accurate, but simply that they are better than no quasi-bids at all. So long as retailers can estimate DR better than AEMO can or could, there will be potential value in implementing the design.

---

[39] That is to say it does not affect the design's functionality. It may affect its effectiveness.

Can Retailers forecast their Base Demand too?

As discussed, the proposed design splits the role of demand forecasting, with AEMO forecasting base demand and retailers estimating and bidding DR. The difficulties of AEMO forecasting DR have been described. However, this leaves open the possibility of assigning the *entire* role to retailers: ie both base demand forecasting and DR bidding. This will be referred to as *retailer self-forecasting*.

Self-forecasting already exists for semi-scheduled generators (SSGs). The maximum amount that they can generate in a DI depends upon local wind or solar conditions and is referred to as the Unconstrained Intermittent Generation Forecast (UIGF). The UIGF is fed into NEMDE and is used in setting dispatch targets.

By default, AEMO forecasts UIGF, but the SSG can choose to self-forecast[40]. Forecast errors will lead to deviations between target and actual generation which, in turn, determine the FPP$ amounts[41]. So accurate self-forecasting will be rewarded with lower FPP charges.

One could envisage a similar model for retailers. But there are some important differences between retailers and SSGs:

- AEMO forecasts UIGFs for each individual SSG[42] whereas its regional demand forecasting is for retailers in *aggregate*; and
- There is real-time metering for individual SSGs, whereas real-time metering of demand only existing at the regional level[43].

Suppose retailers A, B and C operate in a region, and just retailer A wished to self-forecast. AEMO would be left with the task of forecasting the aggregate demand of retailers B and C. But it only has real-time demand information for the aggregate of the three. To forecast for B and C alone, it would need real-time information from retailer A to subtract from the regional demand. Indeed, retailer A would likely itself need real-time information on its customers' demand to be able to forecast more accurately than AEMO.

It would be plausible for a retailer to establish real-time metering on some or all of its customers and self-forecast for these customers, leaving AEMO to forecast the remaining demand in the region. As discussed below, the retailer would likely need to also install 4S metering on these customers, so that an individual deviation can be calculated and processed through the FPP algebra.

At this stage, it seems unlikely that it would be worthwhile a retailer incurring these costs to self-forecast, when AEMO already does this with reasonable accuracy.

Should Visible Demand Response face a specific "Wedge" FPP$ Charge?

As discussed above, the proposed design levies an FPP$ charge on visible DR, based on an expected "wedge" deviation. At face value, this levy would seem to run counter to the objective of encouraging DR visibility. It would be straightforward to exclude such a charge from the design, with the demand deviation instead decomposed into just two components – for invisible DR and noise

---

[40] AEMO assesses the accuracy of these self-forecasts before permitting these to be fed into NEMDE

[41] and also the allocation of regulation FCAS costs

[42] ie each wind or solar farm

[43] and, as already noted, this is in fact based on the proxy of regional generation plus net imports

respectively – and the FPP$ charges allocated accordingly[44]. This section considers the arguments for and against its inclusion.

Chapter 7 explained why and how visible DR still leads to deviations and associated frequency regulation costs. A causer-pays philosophy suggests that retailers with visible DR *should* be charged accordingly.

On the other hand, visible DR brings benefits as well as costs. In particular, the associated improvement in spot pricing brings general allocative efficiency benefits to the market. These positive externalities are real and potentially substantial, but they cannot easily be measured or attributed to individual retailers, and so are not reflected in the design of the incentives. Removing the wedge FPP$ charges might strengthen the incentive for DR visibility and so usefully add to these positive externality benefits.

This begs the question of the materiality of these wedge FPP$ charges. In fact, there are reasons to expect they might be low, or even negative.

The height of the wedge is set by ΔDR: the change in cleared DR between consecutive dispatch intervals. Changes in cleared DR are driven by changes in quasi-bids and/or spot prices. If both are stable, ΔDR will generally be small. Indeed, an objective of making DR visible is that it will lead to greater spot price stability.

Spot price volatility could still arise due to abrupt supply-side changes such as generator outages, but this is in fact likely to lead to *negative* wedge FPP$ charges. To understand this, consider a scenario where a generator outage leads to a supply shortfall and a falling frequency. Although this fall will be quickly arrested by contingency FCAS response, the underlying imbalance remains. This is only corrected at the next dispatch, when additional generation is dispatched and, as a result, spot prices will likely jump higher.

The new generation is urgently needed, to take over from the contingency FCAS response, but the ramped dispatch target means it is not fully dispatched until the end of the DI. On the other hand, the extra DR induced by the spot price jump is provided *immediately*. This step-change response – which leads to the wedge deviation – was previously described as a problem but, in this context, it is a valuable *feature*. The FPP algebra – through which frequency regulation is rewarded – should automatically pay visible DR for this response, through the wedge FPP$ component.

These considerations would suggest that the wedge FPP$ is a worthwhile element of the proposed design.

Should Visible Demand Response enjoy lower Regulation FCAS Costs

As discussed above, the FPP settlement is zero sum: payments to AEMO match payments from AEMO. This means an additional mechanism is needed to recover the costs of regulation FCAS, to replace the pre-existing causer-pays algorithm. The rule change which introduced FPPs includes such a mechanism. Payment amounts calculated under this mechanism are referred to here as FCAS$.

The FCAS$ algebra follows a similar process to the FPP$ algebra: ie:

- Calculate individual supply deviations and an aggregate demand deviation, exactly as in the FPP;

---

[44] Indeed, earlier iterations of the proposed design took exactly this approach

- Correlate these deviations with frequency deviations; and
- Determine FCAS$ amounts based on these correlations.

However, a key difference, is that FCAS$ amounts are only attributed to those with positive correlations[45]: ie those who are creating the need for frequency regulation. FCAS$ amounts are zero for those are *providing* frequency regulation. So with payments *to* AEMO only, settlement is no longer zero sum, and the necessary revenue can be collected by adjusting the price that is applied to the positive correlations.

In principle, the proposed design could allocate FCAS$ between retailers similar to FPP$: that is:

- Decompose the demand deviation into three components, exactly as for FPP;
- Use the FCAS$ algebra to determine three corresponding FCAS$ amounts; and
- Allocate these three amounts as for FPP: ie *wedge FCAS$* in proportion to ΔDR; *flat FCAS$* in proportion to forecast errors; *noise FCAS$* in proportion to retailer demand.

However, there is one potential drawback in this design. Whilst the FPP algebra is linear, the FCAS$ algebra is *non*-linear[46]. This non-linearity means that the decomposition of the demand deviation will lead to an increase in the total FCAS$ cost allocated to retailers. This could be contentious. For this reason, a change to FCAS$ has not been included in the proposed design, but could easily be added in.

How are retailers with 4S metering managed?

Some customers may have 4S metering. This might be the case for large customers, who may have it for their own operational purposes, or in order to offer contingency FCAS. Ideally, 4S-metered customers – and their retailers – could be incentivised to reduce deviations within a DI where this leads to lower frequency regulation costs, similar to how generators are currently incentivised.

Scheduled generators all have 4S metering and have their deviations – and associated FPP$ amounts – calculated individually. Deviations are the difference between reference and actual, with reference based on dispatch targets. That is to say, individual deviations can be calculated because these generators are both 4S-metered *and* dispatched.

Under the proposed design, retailers - even those with 4S metering - would not have dispatch targets. The demand deviation (for all retailers) is based on AEMO's real-time demand forecasts, but retailers do not have those either[47]. Whilst retailer-level demand forecasts are calculated ex-post for settlement, these forecasts represent *average* MW, not the spot MW amounts need to define the reference trajectory.

It might be possible to convert the average MW forecasts into equivalent spot MW forecasts and then calculate individual deviations accordingly. However this possibility has not been explored further. So, under the proposed design, 4S retailers would *not* have individual deviations, unless their load is scheduled or they self-forecast, as discussed above. That means the 4S meter readings would

---

[45] Which are *negative* correlations in AEMO's FPP algebra, because of the opposite sign convention already noted.

[46] In fact, it *has* to be non-linear in order to recover FCAS costs. Linearity would lead to it being zero-sum

[47] unless self-forecasting, as discussed above

not be used in the FPP$ calculations, and the customer/retailer gains no additional financial benefit from this metering.

How does Distribution Congestion affect the Proposed Design?

DNSPs are increasingly seeing the need to manage distribution congestion in real-time. The most likely solution to this uses dynamic operating envelopes (DOEs) that represent real-time distribution network capacity. When congestion emerges, DNSPs are able to directly or indirectly control certain consumer equipment – such as PV or batteries – to ensure network flows remain within the DOE. This is analogous to how AEMO manages transmission congestion today: ie by curtailing generators as necessary.

DOE operation might impact on the amount of DR delivered. For example, consider customer batteries responding to a spot price spike by discharging. If this leads to distribution network congestion, the DOE mechanisms will automatically scale back this response, meaning that less is delivered than expected. That could lead to demand forecasting errors and the associated retailer being penalised in the FPP for "invisible DR".

It is to be hoped that DOE-driven curtailment will be transparent and predictable, allowing retailers to factor it into their quasi-bids: eg if they expect that only 80% of their desired DR response will be delivered as a result of curtailment, they can scale back their DR bids accordingly.

More futuristically, it might become possible for retailers to submit their DR bids to DNSPs, rather than AEMO. The DNSPs could then estimate the level of curtailment and scale-back the bids accordingly before forwarding them to AEMO. There are several new building blocks needed for this to happen, but it is a suitable light-on-the-hill for future development, and the proposed design can be seen as a stepping-stone towards it.

How does Transmission Congestion affect the Proposed Design?

The design proposes regional DR bidding. This is in contrast to bidding of scheduled generation and load, where the exact location of the generation or load must be declared. This allows NEMDE to manage transmission congestion through dispatch: for example, by constraining off generation that is behind a transmission constraint.

NEMDE would not be able to constrain DR in this way, for two reasons. Firstly, because DR is bid at the region level, its precise location is unknown to NEMDE, which would have to assume that it is located "at the regional reference node (RRN)"; that is to say, not behind any constraints.

Secondly, even if NEMDE *were* able to constrain, this would be undesirable as it would create a substantial impediment to DR bidding; given that invisible DR cannot, of course, possibly be constrained. Indeed, it is not clear how even *visible* DR could be constrained in practice in the proposed design, since it does not receive any dispatch instruction and simply responds to changes in the regional spot price.

Currently, AEMO does not generally model or forecast the location of demand, for dispatch at least[48]. In the terminology of the proposed design, the base demand forecasts are regional rather than zonal, so it seems unlikely that zonal DR bidding could usefully help AEMO to manage congestion.

---

[48] This could potentially lead to transmission being overloaded if the load is at a different location, but NEMDE compensates for this through the use of "feedback" constraints, which automatically adjust the transmission constraint to align with actual, metered transmission power flows

If, in the future, AEMO moves to zonal base demand forecasting, zonal DR bids could become useful, and the proposed design could be extended to include this. However, because the FPP incentives only operate at the regional level, there would be no financial incentives for retailers to ensure the zonal DR breakdown was accurate.

Therefore, it seems improbable that zonal visibility of DR could usefully be included in, or promoted by, the proposed design.

Conclusions

The answers to the questions posed are summarised in table 13, below.

Question	Answer
Does the proposed design only affect retailers	It also affects small generation aggregators and, generally, any market participant financially responsible for non-scheduled resources.
Can AEMO's existing regional demand forecasting methods be repurposed	Conceptually, yes. AEMO is best placed to advise on practical and operational issues.
Are all responses of demand to price regarded as demand response?	All responses that AEMO does not model are considered DR.
Can retailers accurately estimate demand response?	Accuracy will not be perfect but will be fit-for-purpose. Retailers will likely specialise.
Can retailers forecast their base demand too?	No. Not without real-time metering data. This is why this is left to AEMO, who has this data.
Should visible DR face a specific FPP$ charge?	Yes. This may incentivise visible DR which helps restore system balance following a generator outage
Should visible DR enjoy lower regulation FCAS costs	Maybe. It would improve incentives but could lead to higher FCAS costs for retailers overall.
How are retailers with 4S metering managed?	Unclear. It may be possible to adapt the design to make use of this data and provide better incentives.
How does distribution congestion affect the proposed design?	DR could be scaled back by DNSPs. Retailers should factor that into their DR bids.
How does transmission congestion affect the proposed design?	DR would be bid regionally and cannot be constrained to help manage congestion.

Table 13: design questions and answers

9. Implementation

Overview

There are two aspects of the design which lend itself to a staged implementation, rather than an all-in-one-go "big bang":

- Much of the new design involves settlement calculations: these can be run in a *shadow mode*, where the new settlement amounts are calculated and published, but existing settlement rules continue to apply to actual transactions;
- Retailer participation is voluntary: retailers can choose the time of entry, in terms of preparing and submitting quasi-bids.

With these aspects in mind, a staged implementation is proposed, as summarised in table 14, below: "current" means in accordance with existing Rules; "new" means based on the design proposed in this report.

Process	**Implementation Stages**			
	Stage 0 Status Quo	**Stage 1 Shadow Operation**	**Stage 2 Part Operation**	**Stage 3 Full Operation**
FPP Actual Settlement	Current	Current	New	New
FPP Shadow settlement	None	New	None	None
Dispatch	Current	Current	New	New
Retailer Quasi-bidding	None	None	Some	Equilibrium

Table 14: Implementation Stages

These stages are discussed in turn below.

Stage 0: Status Quo

This simply involves operating with existing Rules. The FPP Rules and procedures have been designed but not yet developed and implemented. This is due in 2025.

Potentially, the FPP data could be processed at a regional level to decompose the demand deviation – and its associated FPP$ - into the flat and noise components[49], as discussed in chapter 7. A material and growing flat FPP$ could indicate substantial invisible DR, which might prompt a move to the next implementation stage.

Status 1: Shadow Operation

In this stage, shadow settlement amounts (calculated and published but not used in financial settlement) would be determined by running the retailer-level demand forecasting and revised FPP calculations as described in the proposed design. There would be no quasi-bidding or dispatch at this stage, so no cleared DR and no wedge component reflecting visible DR amounts. The

---

[49] recall that the wedge component does not arise until there is quasi-bidding

decomposition would just be between the flat and noise components. Note that this goes beyond the FPP$ analysis in stage 0 by allocating the FPP$ between retailers.

If the algebra works as expected, retailers with DR (which, of course, remains invisible) should see higher shadow FPP charges (based on the new design) than actual FPP charges (based on the current design). These differences might not be material initially, but could be expected to increase over time as the amount of (invisible) DR grows.

Stage 2: Part Operation

In this stage, the design would become fully operational. AEMO would establish the infrastructure to accept and process quasi-bids and, after completion of the qualification stage, input these into NEMDE. Cleared DR amounts would flow through to the new settlement calculations, whose results would now be used for actual settlement.

Possibly, only a few retailers will bid initially, depending upon the materiality of their settlement differences under shadow operation and the impetus this has given them to estimate their DR and prepare their quasi-bidding processes for this stage. So the benefits of better spot price setting might not be fully realised, but at least retailers with invisible DR are no longer pushing costs onto retailers without DR.

Stage 3: Full Operation

Over time, more retailers will begin bidding and the implementation moves to a full operation. “Full” does not mean that every retailer bids. Obviously, those retailers without any DR will have nothing to bid, but even retailers with modest amounts of DR may find it unnecessary or undesirable to bid: ie if the costs of the DR bidding exceed the expected benefits of reduced FPP$ charges.

Also, as some retailers commence bidding, spot pricing accuracy improves, imbalances reduce, and the costs of FCAS – and so of invisible DR – reduce accordingly. Thus it may be that an equilibrium is reached where some but not all retailers are bidding, and FCAS costs remain modest. This represents an ideal trade-off between costs and benefits, a consequence of DR bidding being voluntary rather than mandatory.

Conclusions

A staged implementation helps to align the costs and benefits of introducing the proposed design, with each move to the next stage being predicated on anticipated new benefits exceeding expected extra costs. Thus implementation timing will be aligned with the growth and materiality of demand response.

10. Costs and Benefits

Overview

The various costs and benefits associated with implementing the new design have been discussed above and are brought together in this chapter. This is a qualitative analysis only. Quantification has not been attempted at this stage. Nor is there any attempt to compare costs and benefits with those likely to arise under the AEMO rule change proposal.

Costs arise for AEMO and for retailers. These are discussed in turn below.

AEMO Costs

AEMO costs would include the development, implementation and operation of three new processes:

- Communication, acceptance and processing of quasi-bids,
- Retailer-level demand forecasting, and
- Decomposition of deviations and calculation and allocation of FPP$ amounts for each component.

Some indicative thoughts and considerations are discussed below.

Bidding communications would likely be similar to existing process for scheduled generator and load bids. Acceptance and processing would probably be automated and straightforward. Because the quasi-bids have identical structure to existing (BDU) scheduled bids, there is no anticipated need to change NEMDE functionality. Cleared DR for each retailer is simply based on the “dispatch target” that NEMDE produces for each quasi-bid[50].

The retailer-level demand forecasting is a novel process, albeit based on existing regional demand forecasting methods. As discussed above, it is unclear to what extent the existing method can be migrated to the retailer level and to settlement timescales. So this process is the one where costs are most uncertain.

Decomposition of the deviation follows simple mathematical rules, which should be easy to code and operate. The calculation of the FPP$ amounts uses the existing algebra, so should be just a matter of feeding new inputs into existing software routines. The allocation of these amounts between retailers follows simple rules and, again, development and operation should be straightforward.

Retailer Costs

Costs primarily only arise for those retailers who choose to participate. However, non-participants would still be subject to the new settlement algebra, and there may be some new costs involved in processing and verifying these new amounts.

Those retailers participating face new costs associated with estimating DR within their customer base and then preparing and submitting the quasi-bids.

As discussed, the challenge of estimating DR depends upon the DR business model used. Retailers who control customer load or generation directly through a VPP will likely already have a good idea

---

[50] although, where a retailer offers quantities in fractions of a MW, there would need to be a new process to recover these fractions, since NEMDE operates to whole MW

of the associated DR, like the owner of a real power plant does. However, retailers who send dynamic prices through to customers may not know exactly how customers will respond. Again, there are different models, with some retailers offering load control algorithms for the customer to use, whilst in other cases customers might develop their own algorithms or source these from third parties. These different models provide different levels of challenge for DR estimation.

Bidding would be a new process for standalone retailers (those without generation assets) and may require the establishment of manned “trading desks”. Gentailers will already have these for bidding their generation. However, bidding processes are increasingly automated using “bidding bots”, and such automation could substantially reduce the costs of quasi-bidding.

Benefits

Benefits similarly fall into two categories. There are the direct benefits associated with lower FCAS costs arising from improved DR visibility. And then there are the indirect benefits that the more efficient spot pricing brings to the market generally. These are described in turn.

The design aims to identify the direct benefits arising from each retailer making its DR visible and directly reward retailers accordingly through reduces FPP$ amounts. As already noted, retailers will only incur the costs of DR bidding if these are lower than the expected FPP$ reduction. So the direct benefits should exceed the retailer costs, and the net benefit should always be positive.

The broader benefits of better pricing are more nebulous but would likely be significant nonetheless. Efficient prices lead to allocative efficiency, where producers and consumers make more efficient decisions – in investment and operation – in response to these price signals.

Visible DR is also likely to improve spot price stability, and reduce volatility, by creating elasticity in the demand curve that can help offset inelasticity in supply, particularly when margins are tight and spot prices high. Reduced volatility might also improve price forecasting, particularly over the PD period, and so improve generation and load scheduling.

Finally, visible DR might enhance competition in the spot market and so reduce the impact of market power on prices, again improving pricing efficiency. That obviously depends on who is managing and bidding the DR. If it is simply the major gentailers bidding VPPs, say, market concentration might not change significantly. At the other end of the spectrum, more customers might embrace dynamic pricing – particularly for flexible loads such as EV charging and water heating – putting customers in charge of their demand response, with retailers simply responsible for passing spot prices through to them, and then estimating their DR and conveying that information to AEMO’s dispatch. That is a vision where customer autonomy and sovereignty provide a countervailing power to the concentration of the generation market.

Finally, there will be a general benefit from the lower amount of regulation FCAS needed, leading to lower FCAS prices and releasing valuable assets like batteries to operate instead in the energy market.

Conclusions

Benefits from improved DR visibility will arise from lower costs of frequency regulation and from more efficient, stable and competitive spot prices. The cost savings are allocated to retailers, which will allow them to fund and justify the costs of DR bidding. The benefits of improved spot pricing will be enjoyed by everyone.

AEMO will incur costs from new processes for bidding validation, demand forecasting and FPP settlements. While these costs are kept as low as possible by adapting and repurposing existing processes, rather than creating expensive new ones, it is understood that implementing the RLF is potentially challenging and costly.

Given the costs, there will need to be significant impact from invisible DR – and so associated benefit with making this DR visible – to justify implementation of the proposed design.

11. Overall Conclusions

To conclude:

1. Demand response (DR) is, specifically, the response of non-scheduled resources to short-term changes in the spot price and also, more generally, any similar responses not modelled by AEMO in its demand forecasting algorithms. This is expected to grow over time, as more consumers face spot prices, retailers develop and grow virtual power plants, cheaper and simpler automatic response replaces manual response, and growth in variable renewables leads to more of the spot price volatility that encourages and prompts DR.
2. Under the current market design, increasing DR will lead to growing supply-demand imbalances, and so increasing amounts and costs of the frequency regulation needed to correct these imbalances.
3. DR will also lead to corresponding price-driven variability in metered demand which is likely to adversely affect AEMO's short-term demand forecasting accuracy. This could further exacerbate the imbalances.
4. Changes to the market design will be required to address these anticipated problems. AEMO has submitted a scheduled lite rule change proposal with a "visibility mode" which will help with this. The AEMC has engaged Creative Energy Consulting to develop a design with a similar "visibility" concept, but a different philosophy and approach. This design is described in this report and summarised below.
5. The proposed design addresses and manages the imbalances by incorporating DR into the dispatch process. This means clearing offered generation against a downward-sloping demand curve that incorporates DR; currently, a vertical, inelastic demand curve is used, reflecting the absence of DR at the time that this method was designed.
6. But this new dispatch approach requires estimates of DR amounts and characteristics. These estimates are best made by retailers rather than AEMO. Retailers know their customers and so can build estimates from the bottom up; AEMO would have to use a top-down methodology, which would likely be less accurate and also more contentious, given that these DR estimates would – by design – significantly impact spot price outcomes.
7. DR that is estimated by retailers, and incorporated into the dispatch process becomes "visible", in contrast to the "invisible" DR in the current market. Retailers should be financially incentivised to make their DR visible, rather than this being mandated.
8. DR visibility not only improves dispatch but also allows AEMO to improve its demand forecasting by correcting the demand actuals to remove spot price impacts. AEMO will continue to forecast this price-corrected demand or "base demand". Retailers, being unable to forecast their base demand accurately without the necessary real-time data, are not required to do so.
9. Financial incentives should be proportionate to the value of DR visibility. This can be achieved by basing incentives on the savings in frequency regulation costs associated with this visibility. However, estimating such savings is complicated and is the main task of the proposed design.

10. Frequency regulation costs are to be allocated using a frequency performance payments (FPP) algebra that was introduced in a recent rule change and is due to be implemented in 2025. This algebra uses 4-second (4S) generator metering to precisely calculate the amounts by which generators deviate from dispatch targets. By correlating these deviations with frequency deviations, the impact on frequency regulation can be inferred and costs allocated to generators accordingly.
11. Customers generally don't have 4S metering, so retailer deviations and their associated costs must be inferred from the 5-minute metering data. The FPP algebra in the current rules doesn't attempt to do this, but simply allocates costs in proportion to retailer size. This is a reasonable approximation in the absence of material DR, but fails to distinguish between retailers that have visible DR, invisible DR or no DR. The proposed design adapts and enhances the FPP algebra to make these distinctions and to allocate costs accordingly.
12. The starting point is the insight that invisible DR will lead to larger demand forecasting errors. These are only seen at the region level currently and so cannot be attributed to individual retailers. The proposed design therefore introduces a new "retailer-level demand forecasting process" to determine retailer demand forecasts and forecast errors. This is done ex-post, as part of the settlement process, once customer metering information is available to be aggregated into retailer demand actuals. Existing AEMO forecasting models are repurposed, fed by retailer-level rather than regional demand data, to mimic a real-time forecasting process after the fact.
13. Retailers with and without DR will have different kinds of deviations. Retailers without DR will have demand that simply varies randomly around the forecast and deviations will reflect this "noise". DR, however, will not be random but will be orchestrated by spot price changes. It will create a step-change in retailer demand at the start of each new DI as the spot price changes and DR adjusts accordingly.
14. Based on these insights and inferences, the "residual deviation" calculated by the current FPP algebra, and attributed to retailers in aggregate, can be decomposed into three deviation components, attributable to visible DR, invisible DR and underlying demand variability. Frequency regulation costs caused by each of these components can then be calculated by applying the existing FPP algebra separately to each of the three deviations.
15. These three dollar amounts are then allocated between retailers using different retailer metrics, each reflecting the driver of the respective deviation.
16. Firstly, the visible DR cost component is allocated between retailers in proportion to the change in DR for a retailer between consecutive dispatch intervals. This is because the amount of visible DR is known, but the change in DR nevertheless creates step changes in demand that impacts on frequency regulation.
17. Secondly, the invisible DR cost component is allocated between retailers in proportion to retailer demand forecasting errors. This is because, though the exact amount of invisible DR is unknown, it can be inferred by size of the forecasting errors, based on the insight that invisible DR leads to larger forecasting errors.
18. Thirdly, the cost component arising from underlying demand variability is allocated in proportion to retailer size: ie retail demand. This is similar to the existing FPP design, and reflects the fact that, in the absence of DR, retailer demand variability will impact on frequency regulation costs in proportion to retailer size.

19. With this new settlement algebra in place, retailers with invisible DR will make a larger contribution to frequency regulation costs, other things being equal. They can reduce this contribution by making their DR visible: that is, by estimating the response and submitting these estimates to AEMO, who can then incorporate them into the dispatch process. That, in turn, will lead to a reduction in frequency regulation costs. So the desired incentives for DR visibility have been established.
20. In summary, the proposed design involves four new or modified market processes: the estimation of DR by retailers and submission of these estimates to AEMO; the inclusion of these estimates into AEMO's dispatch and demand forecasting processes; a new retailer-level demand forecasting process, carried out ex-post by AEMO, and used to identify retailers that have invisible DR based on this leading to higher forecasting errors; and a revision to the FPP algebra to break down retailer deviations and associated frequency regulation costs into three components, driven by visible DR, invisible DR and underlying demand variability, respectively. These are allocated accordingly.
21. The proposed design could be implemented in stages. Initially, the new settlement algebra would be implemented by AEMO in a "shadow mode" where the new amounts are calculated and published, but actual financial transactions continue to be based on existing rules. Only once material differences between the outcomes of the old and new settlement methods are seen would the proposed design "go live", with the new bidding, dispatch and settlement processes all operational. Retailers would then individually decide whether and how to estimate and bid their DR.
22. The benefits of the proposed design arise from improved DR visibility leading to lower dispatch costs, lower frequency regulation costs, and more stable and predictable spot prices. Costs arise from the new AEMO processes and also in those retailers who choose to participate. By making participation voluntary, and by largely adapting existing market processes rather than creating brand new ones, the net market benefit is expected to be maximised. This benefit will grow over time as DR grows; the staged implementation approach should help to achieve an optimal timing for the introduction of the new design.

Appendix A: Retailer Models to exploit Demand Flexibility

Overview

Demand flexibility is the willingness of consumers to change their consumption plans (including charging and discharging of batteries, and switching of rooftop PV) at short notice, subject to receiving satisfactory financial rewards from their retailer. Since retailers pay spot prices to AEMO for their customer demand, they have an incentive to encourage customers to reduce, or increase, their demand when spot prices are higher, or lower, than the retail tariff[51], respectively. Alternatively, retailers might just pass on spot prices – or some proxy for them – to the customer, in which case the customer is incentivized directly to manage its demand accordingly.

Different retailing business models will lead to different amounts and types of demand response, and different challenges for retailers in estimating this response. The proposed visibility model does not prescribe or assume any particular business model; rather, it facilitates a diversity of models, with retail competition ultimately selecting the most effective ones. However, to illustrate the likely tasks and challenges retailers will face, a selection of likely business models is presented here:

- Fixed retail tariffs
- Dynamic retail tariffs
- Retailer controls customer loads
- Spot price pass-through

These are discussed in turn below.

Fixed Retail Tariffs

This is the conventional retailing model, currently used to serve the vast majority of customers. As in all the models, the retailer pays the spot price (ie the RRP) in NEM settlement for its customer demand, and recovers the cost of this (and other also other costs such as network charges) through retail tariffs levied on the customer. Fixed does not mean “flat” in the sense of constant in time. Time-of-use (TOU) tariffs which vary by time-of-day etc., according to a schedule specified in the retail contract, are also considered fixed.

Under the definition of demand response used in this report, customers on fixed retail tariffs – and retailers who only use this tariff model – will not have any DR. This is not to say that, in this model, customer demand is static, or passive, or non-responsive to short-term factors. Firstly, if a customer has a time-of-use tariff, they may use their demand flexibility to move demand to lower tariff periods. Secondly, customer demand will still respond to short-term factors such as sunshine (with the PV), temperature (for space heating or cooling) and so on. However, these factors will already be modelled – explicitly or implicitly – by AEMO’s forecasting systems and so will be incorporated into the base demand. So, by our definition, there is no demand response for the retailer to estimate or report. Retailers using solely this business model will not actively participate in the visibility model. That is to say, they won’t submit quasi-bids; however, they will be subject to the same settlement rules as retailers who *do* actively participate.

---

[51] after allowing for network and other charges

Dynamic Retail Tariffs

As its name suggests, a dynamic retail tariff can be changed by the retailer at short notice, typically in response to short-term factors that affect the amount that retailers must pay – to AEMO and to networks – for their customers' electricity use. The most common form of this model currently is a demand management (DM) retail product, under which the retailer may *call* a DM period a certain number of times per year, for a specified length of time and requiring a specified amount of notice. Retailers might do this to manage their exposure to spot prices, or dynamic network tariffs or both.

During a DM period, a customer will typically be required to pay a much higher retail tariff. This might be on its entire load; or it might instead be rewarded at this tariff rate for the amount by which its load is *below* its typical load for a similar period: that typical load level being estimated by the retailer. Either way, the customer has the incentive to use its flexibility to reduce demand over the period. (If it was unable to do this, it would likely not have signed up to the DM tariff in the first place). The retailer has no control over this response, but might estimate it ex-post; indeed, it would need to do this where DM rewards are provided.

It is understood that AEMO is currently unaware of which customers are on the DM tariffs or when DM periods are called, and so does not and cannot factor these into its demand forecasts. So, by the definition above, this gives rise to demand response. However, it is not a response to spot prices: or at least, not to *actual* spot prices. To understand this, consider what happens when a DM period is called. Firstly, the retailer must call this in advance, and so must base the decision to call on *forecast* spot prices (its own forecasts, or those from pre-dispatch). Secondly, once the call is made, the higher tariff is locked in and the customer will respond accordingly, irrespective of *actual* spot prices. Whether spot prices turn out to be $10,000 or $100, the customer response will be the same.

Retailer controls Customer Loads

In this model, a retailer directly controls, in real time, certain customer appliances, batteries or PV. The degree of control will depend upon the technology deployed (ie which appliances receive control signals from the retailer and the form that these signals take) and limitations agreed in the retail contract (eg control is only permitted a certain number of times per year).

A current model of this type is the *virtual power plant* (VPP), where a retailer can control a customer's home battery. Typically, it will do this only a few times a year, to discharge the battery during a spot price spike. The remainder of the time, the battery is instead controlled to reduce the customer's bill: eg by soaking up excess PV output during the day and then discharging during the evening when the fixed retail tariff is higher. The customer may receive a fixed reward (eg $X/year off their bill) and/or a variable reward (eg $Y each time the retailer takes control of the battery).

Retailers will control load to manage exposure to spot prices. Since the customer is not in the loop, a notice period would typically not be needed, and so the response can occur instantaneously or, at least, as fast as the controlled equipment can respond: for example, the home battery might be automatically discharged whenever spot prices exceed $1000/MWh, and this discharge would be signalled as soon as the high spot price number is published by AEMO: ie a few seconds into the DI.

Of course, with batteries and other storage, control ahead of time is useful: eg if high spot prices are forecast for the evening, the retailer might charge a customer's home battery during the day, to be ready to discharge when the spot prices occur. Nevertheless, the actual response is to live spot prices; not forecast spot prices as with the DM model above[52].

The degree of control – and so the accuracy with which demand response can be estimated – will depend upon which consumer loads are being controlled and how. Home batteries can likely be controlled precisely. In extending the model to EV chargers, say, a customer “over-ride” might be included, whereby a customer needing the EV to be charged urgently is permitted to over-ride the retailer’s preference to stop charging during a high spot price period.

Going further, control of a customer’s air conditioning, say, might be through automatic adjustment to the thermostat setting, with the response then depending upon room temperature. Switching the thermostat from 25C to 28C, say, would lead to a response if the room temperature is 26C, say, but no response if it is 23C, or if it is 30C. And, obviously, no response at all if the aircon is not switched on because the customer is not at home. Also, any response is time-limited, since a room starting at 26C will gradually rise to 28C if the thermostat is reset to this level, and at this point the aircon will come back on.

In general, a retailer will know precisely what control signals it is sending to customer appliances, and how these signals change as a function of actual spot prices. However, it may not know exactly the demand response triggered by these signals.

Spot Price Pass-through

Under this business model, the retailer passes on to the customer the cost of the customer’s consumption at the spot price. The retailer is now financially disinterested in the customer’s demand flexibility and has no reason to control the customer’s load.

There are several possible variants of this model. At its simplest, the customer might be left on their own to decide how to use their demand flexibility to manage their retail bill. More likely, the customer will receive some support in the form of hardware (appliance control systems or home energy management systems) and/or software (control algorithms or “apps”). This support could be provided by the retailer or by a third party. It might simply involve some one-off sale of hardware or software for the customer to install and operate. Or, at the other end of the spectrum, the retailer or third party might operate the system for the customer, with the latter’s involvement then limited to deciding on certain settings: eg to require that the EV is fully charged at 7am each morning; or to require that the aircon thermostat is only raised above 25C if the spot price is above $1000/kWh, say.

Knowledge of the app functionality, and possibly the customers’ settings too, would help the retailer to estimate demand response from its customers, even though it has no direct control of load. But, if these services are provided by a third party, the retailer may have no such information[53]. It would be limited to inferring the demand response using a statistical analysis: eg looking at historical correlations between demand, spot price and weather factors.

---

[52] This is not to say that price forecasts won’t play a role. For example, if a retailer is only permitted to control a customer’s battery a certain number of days per year, then it will set the discharge threshold price based on its expectation of spot price volatility for the year

[53] unless it can acquire this from the third party

Conclusions

Demand response is defined as the response of customer demand to short-term factors which are not used or reflected in AEMO's demand forecasts. It will arise in retailer business models which harness customers' demand flexibility to respond to these factors. Depending upon the model, retailers may have more or less control of, or information around, this response, which will affect their ability to accurately estimate this response.

Appendix B: Changes to the Design in 2024

Overview

A version of this report was prepared by CEC and published by the AEMC in December 2023 under the title of *A Scheduled Lite design to integrated Demand Response into NEM Pricing and Dispatch*. CEC was subsequently engaged again by the AEMC, and some changes to the proposed "2023" design were discussed and agreed during this engagement, to address issues raised by the AEMC and by stakeholders. These design changes are incorporated into this second version of the report.

This appendix lists and explains the main changes between the *2023 design*, described in the first version of the report and the *2024 design*, described in this report.

Base Demand and Demand Response

This is more of a definitional issue than a design change. The 2023 report defined demand response as a response of non-scheduled resources to spot prices. This meaning has been generalized somewhat, to cover any kind of response that is not modelled by AEMO in its short-term (dispatch and pre-dispatch) demand forecasting algorithms. This might, for example, include response to dynamic network tariffs, unless and until AEMO incorporates that response into its forecasting processes.

This definition makes it vital that AEMO's forecasting process is transparent, so retailers know exactly what kinds of response are, or are not, included.

Structure of Quasi-bids

The 2023 design was non-specific about the structure of quasi-bids, and noted some possible design options and variations. The 2024 design uses an identical structure to the bid structure recently introduced for scheduled bi-directional units (BDUs). This allows the quasi-bid to incorporate both positive and negative values of DR, at different prices. The only minor difference from the BDU bid structure is that some decimal places might be permitted in the offered quantities, whilst BDU bids must use whole MW quantities.

Bid Aggregation

The 2023 design proposed that AEMO aggregate all of the quasi-bids within a region into a single aggregate bid that was submitted to NEMDE. However, given that the quasi-bids now take the BDU structure – and so are able to be submitted into NEMD directly – such aggregation is now complex and unnecessary and so is removed from the 2024 design. Instead, each quasi-bid will be submitted directly into NEMDE, broadly equivalent to having a separate DUID for each bidding retailer.

Timing of Quasi-Bids and Rebids

The 2023 design noted the need for quasi-bids to be submitted in time for dispatch, but was ambivalent about whether earlier submissions would be required or desirable. The 2024 design specifies that quasi-bids would need to be submitted day-ahead, to the same timetable as that for scheduled bids. This is to ensure that the DR contained in quasi-bids is reflected in pre-dispatch outcomes. As with scheduled bids, the offered quantities in quasi-bids would be able to be re-bid closer to dispatch, but not the offered prices.

Retailer Qualification and Disqualification

This is a new process introduced into the 2024 design. When a retailer commences quasi-bidding for the first time, its bids would not initially be input into NEMDE. The retailer would first have to go through a qualification period, during which the accuracy of these quasi-bids would be measured and evaluated. This would be done by running the retailer-level forecasting process – for that retailer – with and without the quasi-bids included[54] and then comparing the accuracies of the two sets of forecasts. The qualification period would be long enough to give AEMO comfort that the quasi-bids would improve forecasting performance (compared to forecasting without using the quasi-bids) in this way under a full range of NEM conditions.

A retailer could later be disqualified if AEMO found that the quasi-bids were adversely impacting dispatch efficiency or frequency performance, and would then need to re-qualify.

Regulation of Quasi-bidding

The 2023 design considered the question as to whether the AER should monitor quasi-bidding to identify false or misleading quasi-bids. This is now a confirmed element of the 2024 design. In addition, the AER would be able to disqualify a retailer it found to have submitted such bids.

---

[54] recall that the quasi-bids lead to cleared DR which is used to price-correct the retailer metered demands that are fed into the forecasting algorithms