A Scheduled Lite design to integrate Demand Response into NEM Pricing and Dispatch

Prepared by

Creative Energy Consulting

December 2023

Executive Summary

Background and Context

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 stable and predictable. AEMO has been able to forecast the power demand created by them 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) will lead to growing supply-demand imbalances in dispatch, and so 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 has 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.

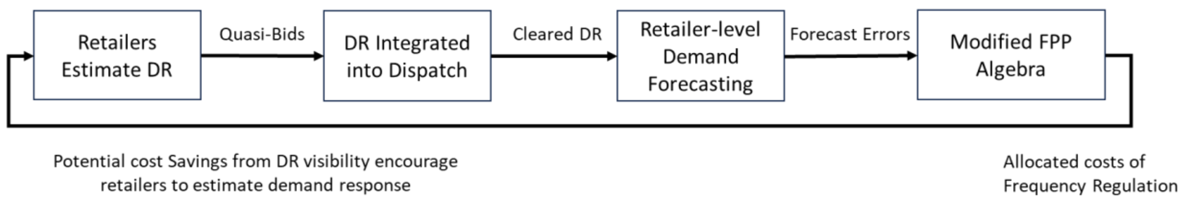

Figure E1: Overview of Proposed Design

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 demand response, the demand curve is in reality 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 demand response will lead to growing imbalance between supply and demand.

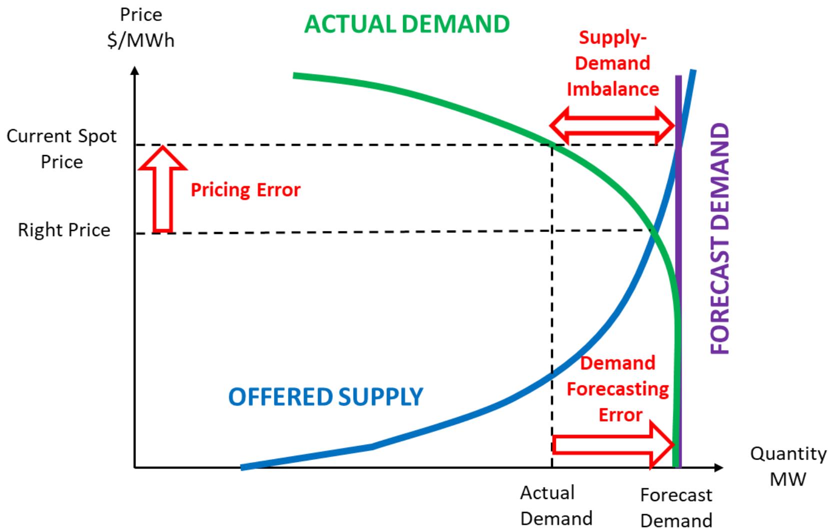

Figure E2: Integrating DR into Dispatch

Demand response 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. The proposal is for retailers to each estimate their own DR and submit these estimates to AEMO in the form of "quasi-bids". This is preferred to AEMO making its own estimates of demand response. 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, given that these DR estimates would – by design – significantly impact spot price outcomes.

DR that is estimated by retailers, and 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". 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 quickly lead to 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. This report specifies how to adapt and enhance 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 new design proposed in this paper 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 forecasting models, fed by retailer-level rather than regional demand data. Although done ex-post, the process mimics the real-time operational process and is not allowed to "cheat" using the benefit of hindsight.

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 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 "residual 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.

Costs and Benefits

The proposed design involves four new or modified market processes:

1. *DR 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 or not retailers actually participate. 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 replace or complement AEMO's rule change proposal	It would enhance AEMO's visibility mode and could run in parallel with AEMO's dispatch mode
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?	No, only responses to spot prices; not responses to other prices such as retail tariffs.
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.
Can quasi-bids be safely input to NEMDE?	This will occur only once AEMO is confident that they are reliable and accurate.
Might retailers be incentivised to falsely bid DR in order to manipulate spot prices?	Possibly. Regulations should be introduced to prohibit this behaviour.
What timescale should DR bids cover?	In dispatch, to reduce frequency regulation; and possibly in pre-dispatch to improve scheduling.
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.
Does it matter that frequency regulation generally covers multiple regions?	No. The design can and will accommodate this in the settlement algebra.

Table of Contents

1. Introduction and Background

AEMO Rule Change Proposal

AEMO has 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 has 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;
- *AEMO design* or *AEMO proposal* refers to the AEMO scheduled lite rule change proposal[1];
- *Proposed design* refers to our proposed design as set out in this paper.

Report Structure

This report is structured as follows:

- Chapter 2 describes how and why growing spot 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 4-6 describe the key elements of the design in more detail, their underlying rationale and expected outcomes;
- Chapter 7 aims to anticipate and answer possible stakeholder questions around the design rationale, details and possible extensions or amendments;
- Chapter 8 describes a possible implementation strategy;
- Chapter 9 qualitatively assesses likely costs and benefits.

[1] https://www.aemc.gov.au/sites/default/files/2023-01/ERC0352_Rule%20Change%20Request_Scheduled%20Lite%20-%20including%20Appendix.pdf

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	A change in demand as a direct or indirect response to a change in the spot price. 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 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]
DI	Dispatch Interval
DR	Demand response
FCAS$	Regulation FCAS cost allocation
FPP	Frequency Performance Payments
FPP$	FPP settlement amount
LSU	Light Scheduling Unit
NEMDE	NEM Dispatch Engine
PASA	Projected Assessment of System Adequacy
PD	Pre-dispatch
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[2] 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 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[3].

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] mechanisms through which these signals are provided are discussed in chapter 7

[3] 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.

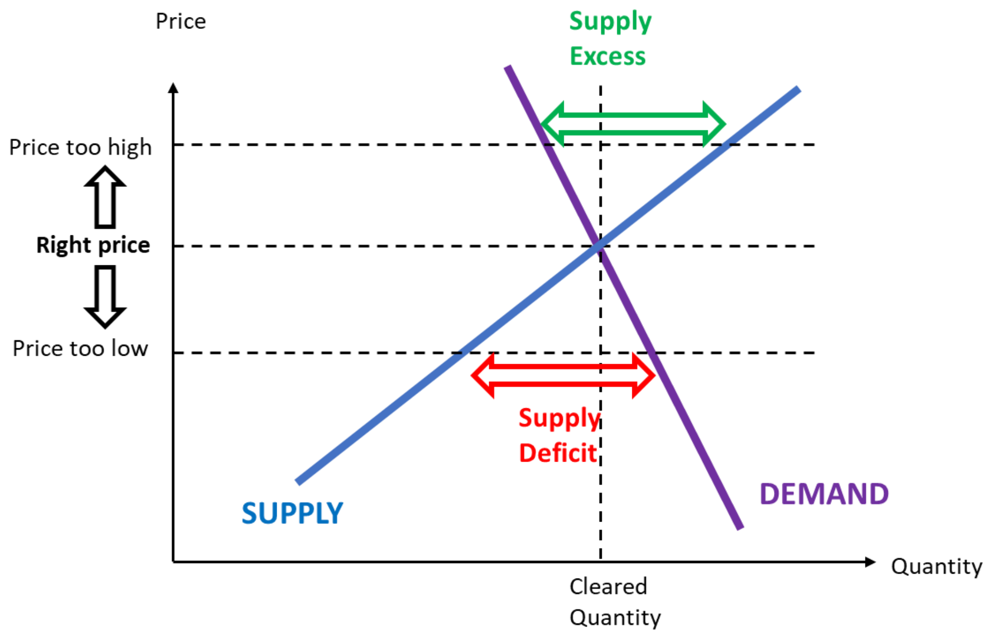

Figure 1: The right price ensures supply and demand are in balance

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[4], 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[5]. Since any resource can either be a generator (producing electricity) or a load (consuming electricity), resources fall into 4 categories[6], 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

[4] 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.
[5] 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.
[6] 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[7]; 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.

[7] 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 is 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.

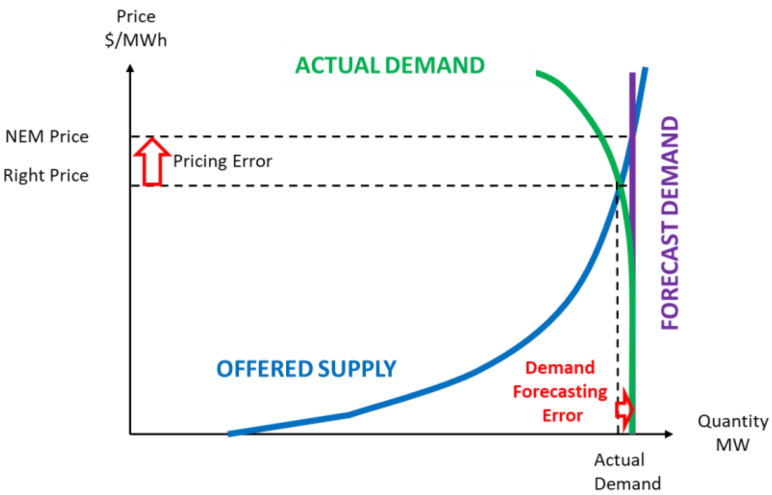

Figure 2: with inelastic demand, price and demand errors are small today

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

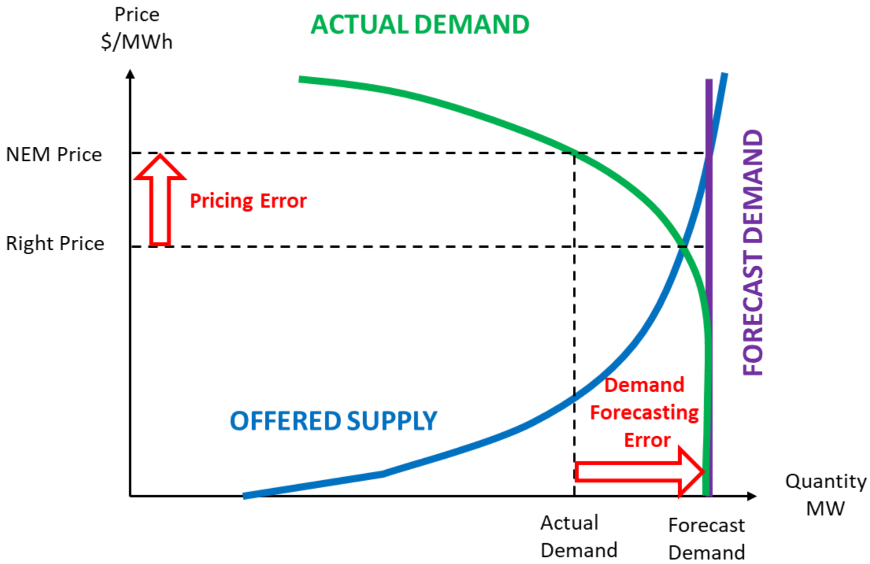

Figure 3: with elastic demand, price and demand errors would be much larger

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[8].

This is illustrated in figure 4, below.

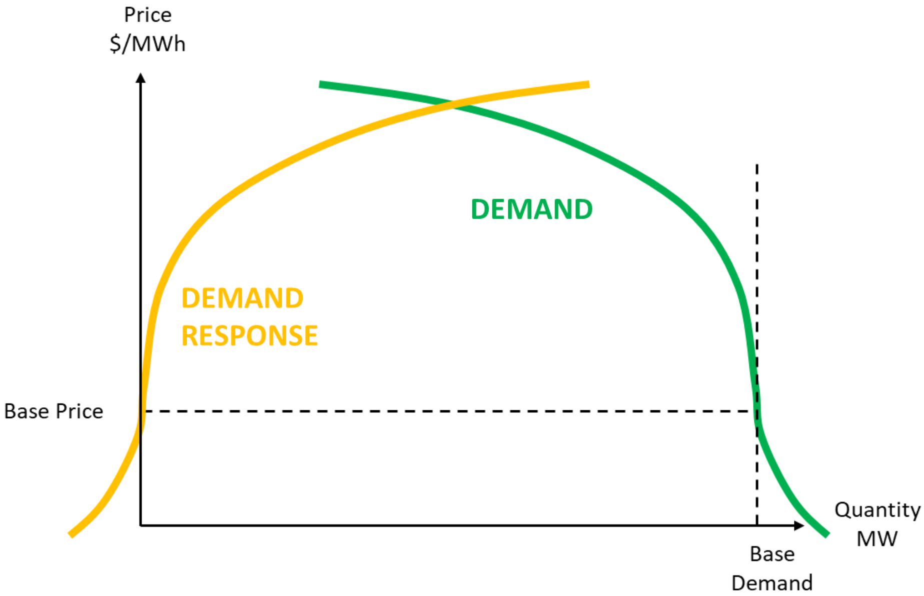

Figure 4: decomposition of demand into base demand and demand response

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.

[8] 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	0	10
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 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

Table 5: types of frequency regulation

Frequency Performance Payments (FPP) were introduced through a recent rule change and are due to be implemented in 2025. As discussed below, they are used in the proposed design to provide the 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

Demand response 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, 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[10]. 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.

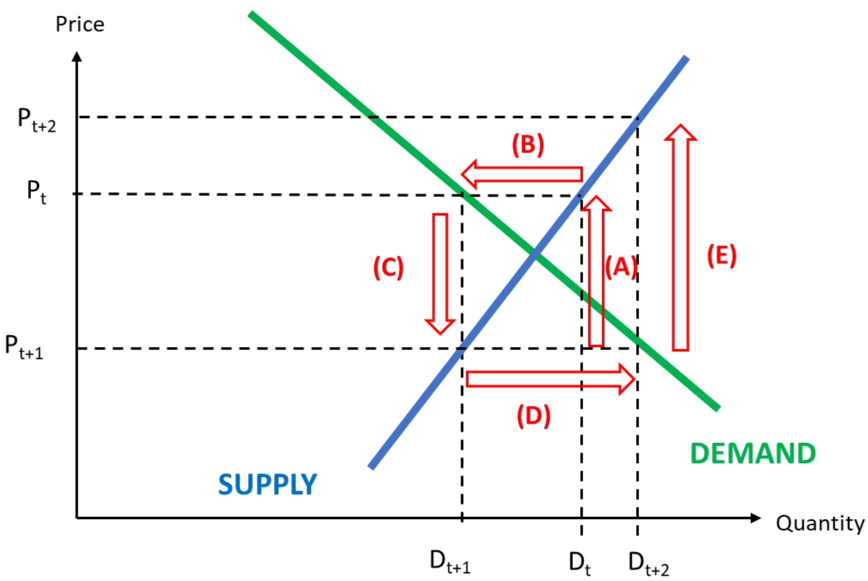

Figure 5: A hog cycle in Dispatch

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

[10] this would be based on the actual demand for the prior dispatch interval

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[11]. 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. The fundamental challenge[12] of the visibility reform being considered in this rule change project 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.

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

[12] that is not to say it is the *only* challenge. Other challenges arising in relation to unpredictability or uncontrollability of non-scheduled resources are considered in Chapter 7

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 whilst retailers estimate demand response;
2. Retailers are financially incentivised to provide accurate estimates of demand response 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 likely stakeholder acceptance; and
4. Obligations on retailers that choose to estimate DR remain light[13] 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

Forecasting base demand is assigned to AEMO because short-term demand forecasting relies on knowledge of current demand. AEMO has this information through SCADA metering[14]; retailers do not, because customer meter readings are generally not collected until the following day[15]; even if they were available in real-time, aggregation would be a difficult and time-consuming process.

On the other hand, estimating 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 7.

Incentivisation

Financial incentivization encourages an efficient trade-off between the costs of accurate DR estimation and the benefits; at least if the incentives reflect the benefits, which is the intent in this design. Some retailers may have little or no DR; others might find it very difficult and expensive to

[13] "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.

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

[15] and later for non-smart meters

estimate the DR they have. In each case, 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 restructuring them as schedule load 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.

It is hoped that this approach will 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[16]. Obligations proposed for DR are compared with those existing for scheduled load in table 6 below.

	Scheduled Load	**Scheduled Lite (visible DR)**
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.

[16] 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.

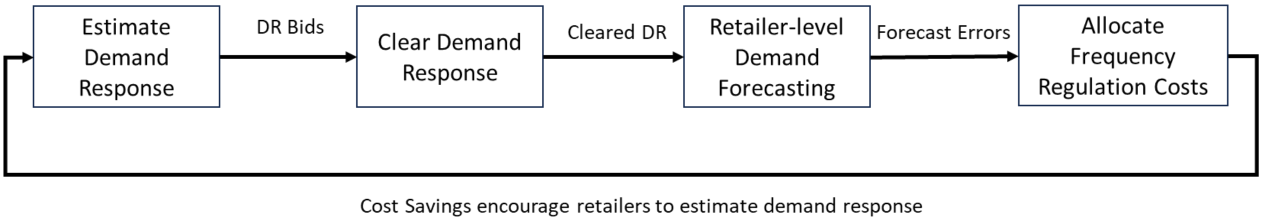

Figure 6: Four Steps are used to create incentives for visibility of demand response

The key steps are as follows:

- *Step 1: Estimate DR*: retailers can estimate their customer DR and submit this information to AEMO;
- *Step 2: Clear DR*: DR information is fed into NEMDE, as scheduled load 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.

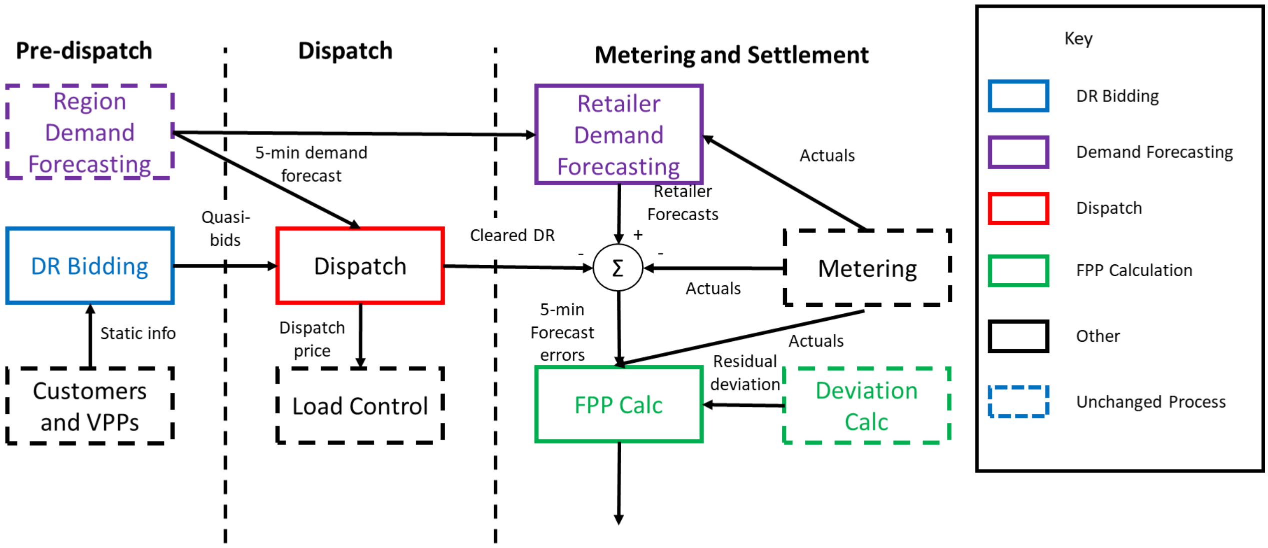

Figure 7: high-level architecture of Proposed design

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

- Chapter 4 describes the estimation, bidding and dispatch of demand response;
- Chapter 5 describes the new *retailer-level demand forecasting* process;
- Chapter 6 describes changes and extensions to the FPP process.

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, Bidding and Dispatch

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. The proposed design addresses these problems by feeding retailer DR bids into the dispatch process and by then using cleared DR amounts to price-correct the demand history. The process architecture for bidding and dispatch is shown in figure 8, below. The regional demand forecasting architecture is shown in a figure 10, further below.

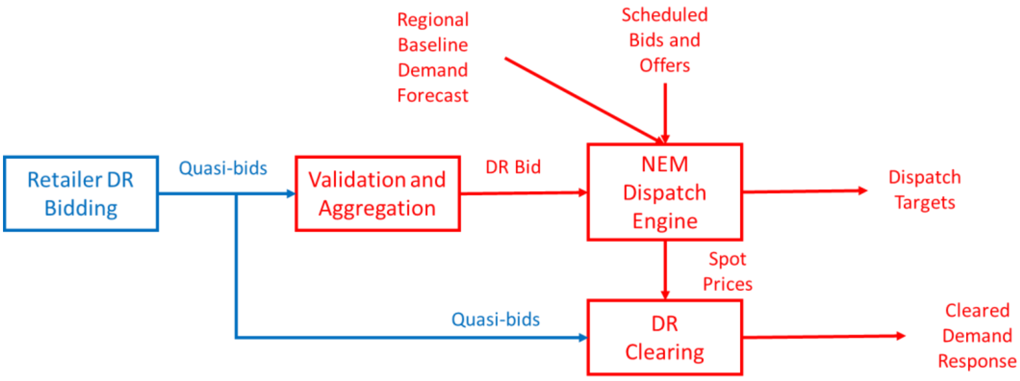

Figure 8: Detailed Architecture for Bidding and Dispatch Processes

Retailers estimate their DR and make this visible by submitting *quasi-bids* to AEMO which describe this estimated response curve. AEMO validates and aggregates these bids and submits this aggregate bid to the NEM dispatch engine where, in combination with the base demand forecast and generators bids, the market is cleared to determine spot prices, dispatch targets and *cleared demand response*. The latter is then used to price-correct actual meter readings, to infer the *actual base demand*.

Retailer DR Bidding

The DR information submitted to AEMO by retailers has obvious similarities to scheduled load bids, but there are some important differences. Firstly, retailer demand – unlike scheduled load – is *not* dispatched by AEMO; there are no dispatch targets which retailers must conform to. Secondly, the retailer bid does not carry any information on the base demand, only the demand response: movements away from the base demand in response to movements of spot prices away from the base price. To highlight these important distinctions, the DR submissions will be referred to as *quasi-bids*: they look like bids but are *not* bids.

Conceptually, the demand curve – and so the demand response curve - is continuous and must be described at every price point. For reasons of practicality and proportionality, the number of points described will necessarily be limited, as shown in figure 9, below.

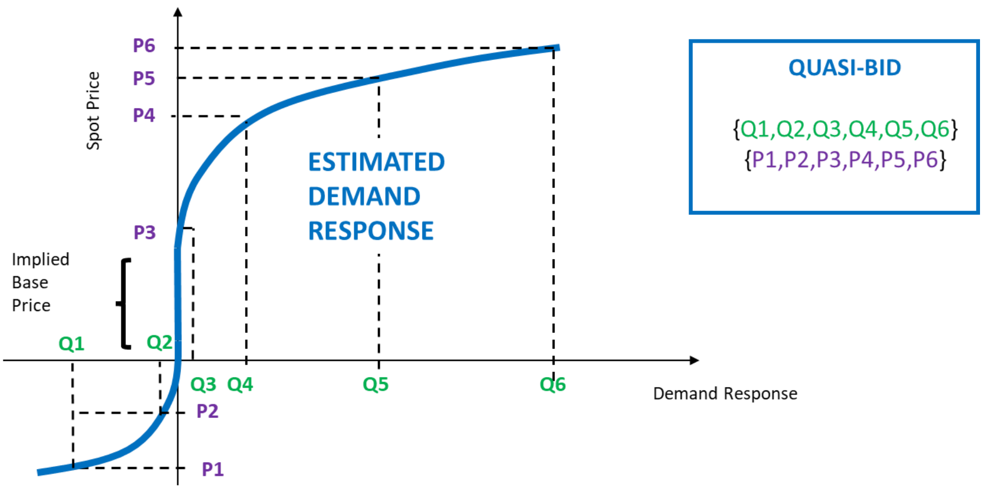

Figure 9: converting a continuous demand curve into a discrete quasi-bid

Table 7 provides a numerical example of a quasi-bid.

Price Point ($/MWh)	Demand Response MW
-1000	-200
-100	-100
0	0
50	0
100	100
500	200
1000	500
15000	1000

Table 7: Example of a quasi-bid of demand response

Note that:

- DR will be positive (ie demand less than base demand) for prices above the base price and may be negative for prices below the base price.
- The base price chosen by the retailer is implied – within a range (eg in table 7 it is between $0 and $50). The Rules need not specify the base price that retailers should use, although they could. It is straightforward to change the base price by adding a constant amount to all the quantities: eg adding 100MW to the quantities in table 7 would move the base price to -$100.
- The quasi-bid does not describe how DR varies between the price points, and the detailed design will need to specify how AEMO interpolates between these points.

- Retailers will be incentivised to ensure their bid covers the full range of spot prices over which its DR is expected to vary. In the example it is from the market floor price to the market price cap, although this would not always be the case.
- There is no particular reason[17] to limit the quasi-bids to 10 price points, as is the case for actual bids. That number was chosen at NEM commencement, primarily reflecting the price response of thermal generators and the IT costs associated with handling and processing bids.
- There is also no reason for these prices to be fixed day-ahead. Allowing them to vary could allow retailers to better represented their DR estimates in their quasi-bids.
- Finally, there is no reason for quantities to be in whole MW. Indeed, this limitation could provide insufficient granularity for very small retailers, whose DR might even be less than 1MW in total.

DR quasi-bids are incorporated into NEMDE and so must be submitted in time for this. However, there may be value in requiring *earlier* bidding, for use in pre-dispatch (PD) or even the Projected Assessment of System Adequacy (PASA), discussed further in chapter 7. There would need to be a reliable and fast IT system to transmit retailer bids to AEMO. Whilst the existing system used for generator bidding could be potentially be used and extended, it might be possible to instead offer a "lite" system with lower costs for smaller retailers to use[18].

Bid Validation and Aggregation

Because the quasi-bids will affect dispatch outcomes, AEMO would need to undertake some bid validation. This might be as simple as looking for gross "fat finger" errors, or it might compare the bids to previous bids made under similar conditions (eg at the same time of day in previous weeks).

The next step would be to aggregate and restructure the bids. Since retailer-level dispatch targets are not required, NEMDE only needs to know the *aggregate* DR in a region, not how much is contributed by each retailer.

To avoid having to make changes to NEMDE, the aggregate DR bids would be presented to NEMDE in the same form as scheduled load bids: that is with 10 offer bands with associated prices and quantities. If 10 bands were insufficient to accurately represent the aggregate DR bids, two or more scheduled load bids could be created, providing for multiples of 10 offer bands.

This processing would be done immediately prior to dispatch, to allow for the most up-to-date quasi-bids or from retailers.

It is critical that the quasi-bids fed into NEMDE help to improve dispatch and reduce supply-demand imbalances, or at the very least do not exacerbate these. This is discussed further in chapter 7.

[17] that is to say, no *conceptual* reason. There may be practical reasons; for example if existing systems used to exchange or process bids are re-purposed for quasi-bidding

[18] indeed, for a very small retailer, whose DR quantity is too small to impact dispatch outcomes, it doesn't actually matter whether it is reliably received prior to dispatch, so long as it is in place in time for the settlement calculations

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 10.

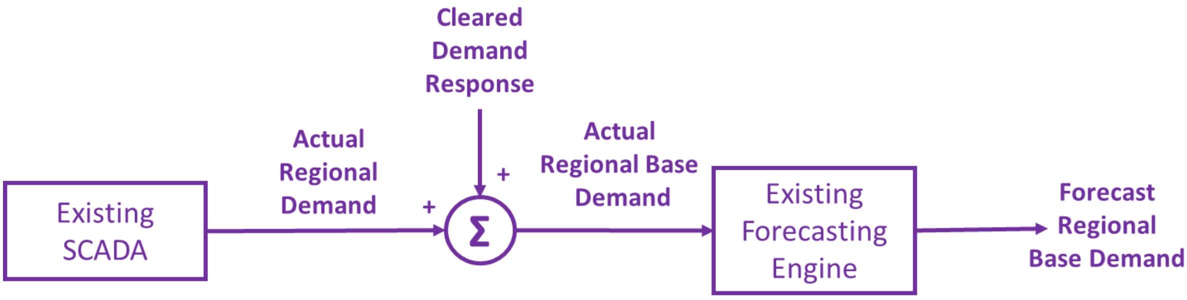

Figure 10: detailed architecture of regional base demand forecasting

Immediately after dispatch, when spot prices have been determined, AEMO will calculate the amount of *cleared DR* in each region in each DI; that is to say, the aggregate amount of the scheduled load bids representing DR that are in-merit at the actual spot price. The cleared DR amount is added to the actual (metered) demand to determine the price-corrected base demand.

The calculated base demand is now predicated on the base price used in the 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[19] 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.

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.

[19] perhaps triggered by a major generation outage

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.

Dispatch

NEMDE would, as now, calculate dispatch targets and spot prices. Inputs to NEMDE would be the same as today, except for the following changes:

- The regional demand forecast used is now a forecast of *base* demand as describe above;
- New scheduled load bids, representing the quasi-bids submitted by retailers, would be included.

Dispatch targets would be sent to scheduled generation and load as now. However, there would be no need for DR dispatch targets to be sent to retailers, for several reasons:

- Since incentives are in place for DR quasi-bids to be accurate, there is no reason for retailers (or their customers) to do anything different to what they said they would do. They will see the published spot price and respond accordingly[20];
- NEMDE can calculate cleared DR for each retailer, but not their base demands; AEMO forecasts this, but only at the regional level, not the retailer level[21]. Thus, AEMO would be unable to calculate demand dispatch targets for retailers to follow; and
- There is no real-time metering of non-scheduled load, so AEMO could not see any deviation from dispatch target, even if there were a target for the retailer to follow.

Thus, the "dispatch" of DR is solely to ensure that the right amount of scheduled generation and load is dispatched and that the spot price is correct. In this respect, it is really an adjustment to the dispatch algorithm rather than a dispatch of non-scheduled load, *per se.*

Conclusions

The task of estimating the demand curve is split between AEMO and retailers, with AEMO estimating the base demand and retailers estimating the demand response: the variation from this base level in response to spot price changes. This division of labour allows each to make the best use of information already available to them: AEMO knows the base demand – at the regional level – in real-time, whilst retailers know about their customers and how they will respond to price. This avoids the need for either party to have new information, which could be difficult, costly or contentious.

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.

[20] note that, unlike scheduled load, DR cannot be constrained on or off by AEMO, discussed further in chapter 7

[21] retailer level base demand forecasts are generated later, as discussed in chapter 5, but this is too late to be used for dispatch conformance monitoring

5. Retailer-level Demand Forecasting

Overview

Retailer-level demand forecasting 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[22];
- 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 11.

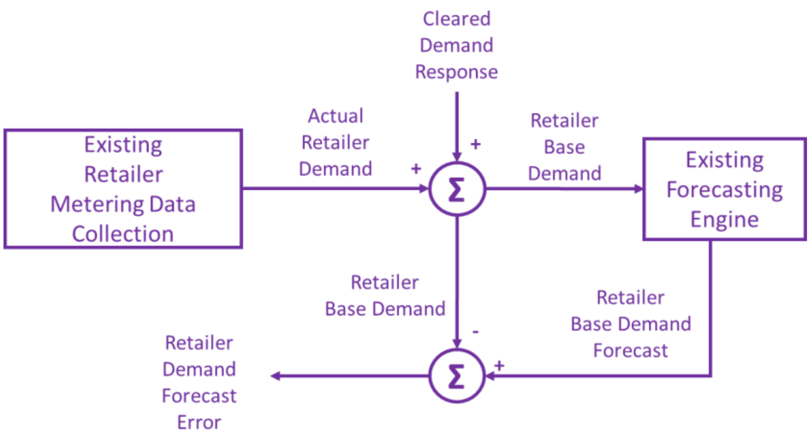

Figure 11: detailed Architecture of Retailer-level Demand Forecasting

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.

[22] 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 is a detailed design issue which is not considered further here. It is assumed that the necessary data will be available for the new process described.

Not used operationally

Demand forecasting is key to system operation in general and dispatch in particular. So why are these retailer-level forecasts not used operationally? The answer is that this does not give any useful additional information to dispatch that is not already contained in the regional demand forecast. Potentially, if the forecasts were bottom-up – provided by retailers and then aggregated – there might be some new information. But this is not happening here, and cannot realistically happen unless and until there is real-time information on retailer demand[23].

Rather, the 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 completely and accurately represented in quasi-bids.

Undertaken by AEMO

Since the forecasts are used in determining settlement payments, it is important that they are done by AEMO to ensure objectivity and auditability. Of course, if it were left to retailers, they would be inclined to "cheat" by using information on actuals.

Carried out after real-time

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, 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.

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. No cheating allowed!

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.

Using existing AEMO forecasting models

The term "models" rather than "systems" is used, because the operational and settlement contexts are very different and likely to be run on different platforms by different staff. But, ideally, the models should be the same; or at least made as similar as practicable. This saves on the cost – and controversy – associated with designing and building new models.

Will the models be suited to this new task? 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

[23] this is discussed further in chapter 7

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 by the model is available at the regional level but not the retailer level. One possible problematic variable is rooftop PV output, discussed further in chapter 7.

Price-correction

As already noted, demand actuals will include any DR and must be price-corrected by adding back the cleared DR. This is calculated for each retailer in each DI by finding the quasi-bid that was active in that DI and then reading off the quantity of DR bid at the outturn spot price[24].

This corrects the demand back to the base demand at the base price chosen by the retailer. So long as the retailer is consistent in its choice of base price, this will lead to a stable, price-corrected demand series.

Forecasting Errors

The forecasting error, for each retailer, is then simply the difference between the forecast and actual base retailer demands in each DI.

To understand the relevance of forecasting errors, consider the impacts of *visible DR* and *invisible DR*:

- Visible DR is DR that a retailer estimates and submits in a quasi-bid.
- Invisible DR is DR that a retailer does not estimate or bid; or bids inaccurately.

It was discussed in chapter 2, how DR is invisible under the existing market design, and this is liable to give rise to demand forecasting errors, because AEMO's model does not and cannot incorporate DR. There will be a similar impact at a retailer level: a retailer with invisible DR will, other things being equal, have larger forecasting errors than a retailer with visible DR. So, a retailer can reduce its forecasting error by estimating and bidding its DR.

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.

[24] this might require some interpolation, where the spot price falls between two price points in the quasi bid

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[25].

It will be seen in chapter 6, 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[26]. Therefore, small retailers will not be unfairly penalised as a result of having proportionately higher unsystematic forecast errors. Such errors will be settlement neutral.

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[27].

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] 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

[26] where charges are allocated in proportion to retailer size, as discussed in chapter 6

[27] 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

6. 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[28]. 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 recently developed (but not yet implemented) frequency performance payments (FPP) scheme[29]. 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 12, below. The main processes are described in the following sections below.

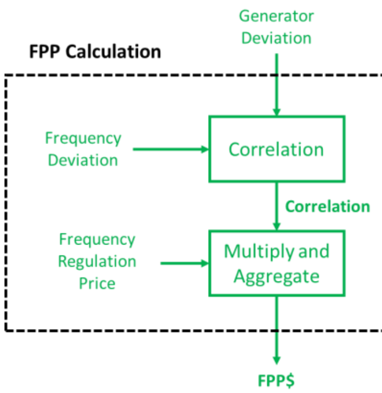

Figure 12: Architecture of the Current FPP Scheme

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

[29] introduced in AEMC's Primary Frequency Response Incentive Arrangements final rule

Generator Deviations

Generators are scheduled[30] 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), at 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 4S deviations – the difference between reference and actual – to be calculated, as shown in figure 13 below.

Figure 13: Calculating Generator Deviations (taken from AEMO’s FPP Procedure)

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[31].

[30] 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.

[31] 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

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[32]:

- 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

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:

[32] 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.

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*[33], 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[34]. 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.

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.

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

[34] 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.

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:

Residual 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 14, below.

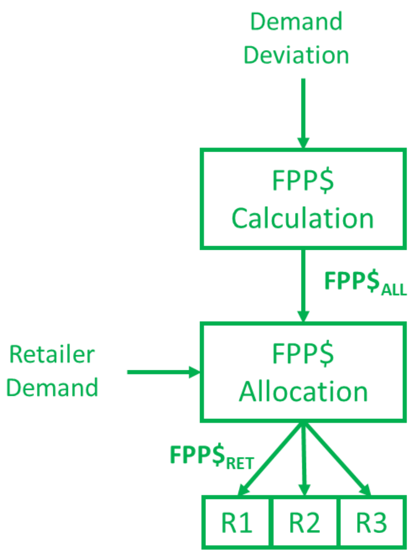

Figure 14: Current Architecture for calculation of retailer FPP Amounts

A simple example of this allocation process is presented in table 8, 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 8: 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 15, below.

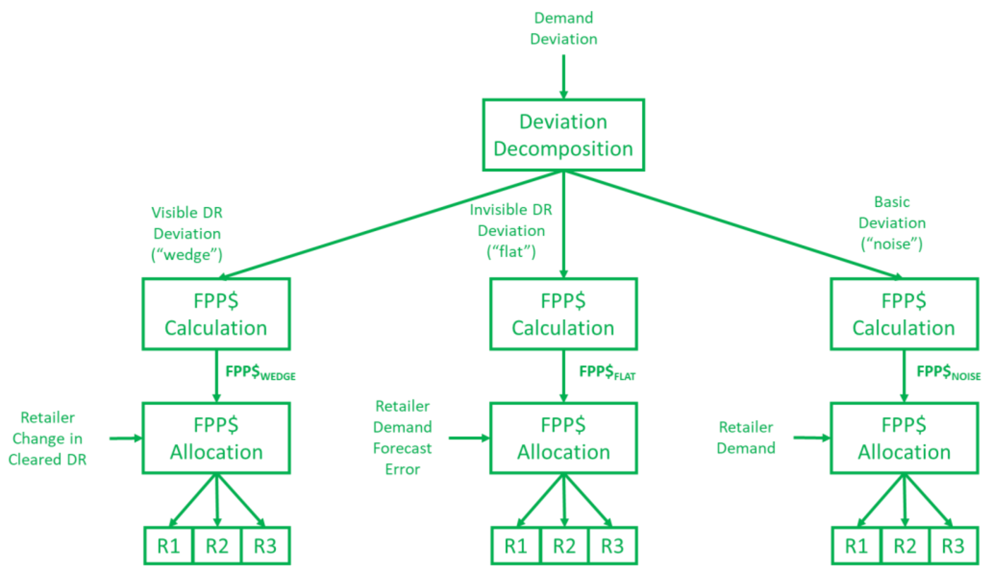

Figure 15: Revised Retail FPP$ Architecture in Proposed design

Decomposing the deviation

The proposed design decomposes the demand deviation into three components, as shown in table 9, 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 9: the three Components of the Demand Deviation

Deviation from Invisible Demand Response
Consider a simple scenario, illustrated in figure 16, 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.

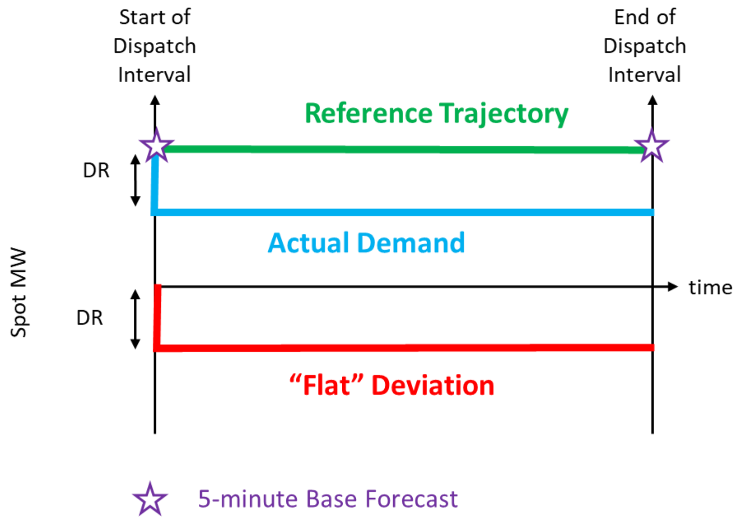

Figure 16: Invisible Demand Response creates a Flat Deviation

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 17 below.

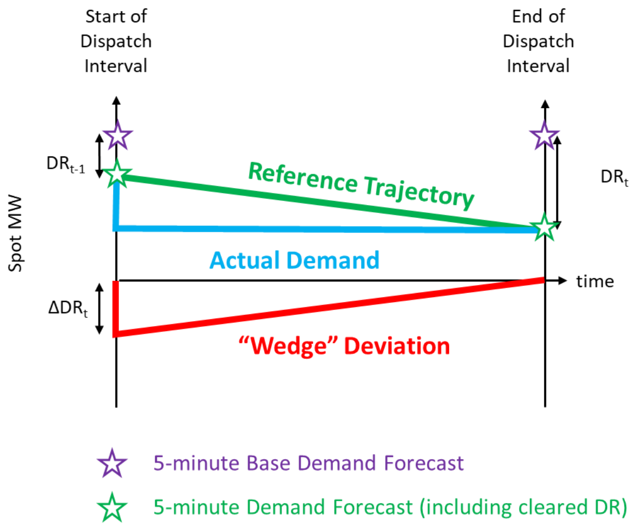

Figure 17: Visible Demand Response creates a Wedge Deviation

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 18, below.

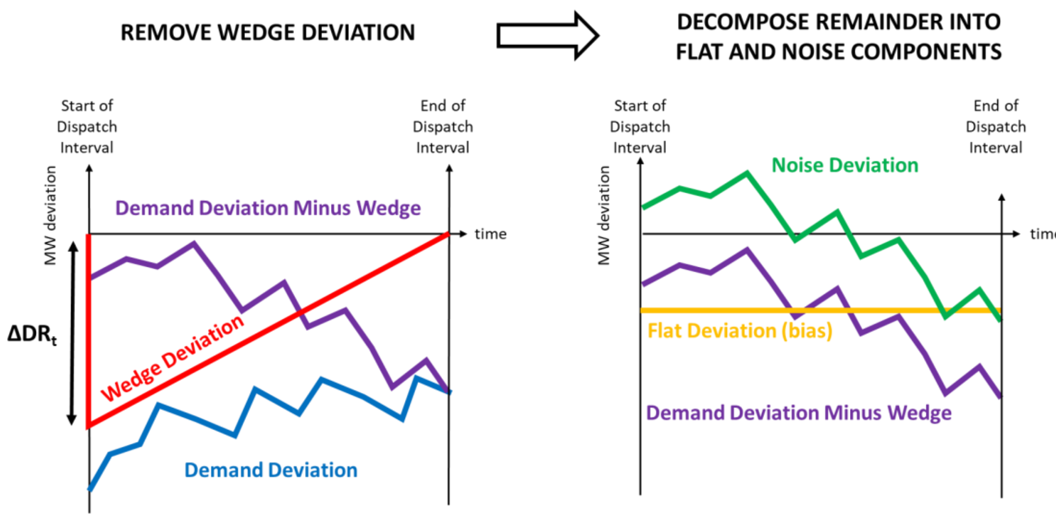

Figure 18: Decomposing the Demand Deviation

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 18 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[35].

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 18, this is just a random deviation with zero bias.

[35] 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 10 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 10: 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[36]. 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[37], 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.

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

[37] 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 5 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[38].

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 5, 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.

[38] 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 7

Numerical Example

A numerical settlement example is shown in table 11 below. This takes the \$2000 $FPP\$_{ALL}$ amount from the previous example shown in table 8 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 11: 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 11). 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.

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.

7. Design Assessment and Discussion

Overview

The proposed design has been described and discussed over the previous three chapters. This chapter raises and discusses some potential concerns or issues that might be raised by stakeholders. It also discussed 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 replace or complement AEMO's rule change proposal?
- 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?
- Can quasi-bids be safely input to NEMDE?
- Might retailers be incentivised to falsely bid DR in order to manipulate spot prices?
- What timescale should DR bids cover?
- 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?
- Does it matter that frequency regulation generally covers multiple regions?

These questions are considered in turn below.

Does the proposed design replace or complement the AEMO rule change proposal?

This report does not aim to describe or critique the AEMO rule change proposal. Nevertheless, it is useful to consider whether the proposed design addresses the issues that AEMO is concerned with, and to what extent.

AEMO proposed that a defined portion of a retailer's customer base could be carved out into a separate light scheduling unit (LSU). An LSU could participate in one of two modes:

- *Visibility mode*: the retailer would provide visibility of DR within the LSU
- *Dispatch mode*: the retailer would bid the LSU similarly to a scheduled load and AEMO would dispatch it accordingly.

The proposed design is essentially an enhancement to AEMO's visibility mode: it provides the DR visibility that AEMO is seeking, and also provides a mechanism for calculating and calibrating the incentives that AEMO's visibility model requires. It is not an alternative to the dispatch mode, because the DR is not dispatched or subject to the associated compliance requirements. But it could potentially complement dispatch mode, with the proposed design and AEMO's proposed LSU dispatch mode operating side by side.

Under this "hybrid" model, a retailer could opt to establish a dispatchable LSU under the AEMO design. To avoid overlap and duplication, this LSU – and the associated demand – would be excluded from the proposed design. A retailer's quasi-bids would only cover its *non*-LSU demand response.

Correspondingly, AEMO's five-minute forecasting would be exclusive of LSU demand, both in real-time[39] (at the regional level) and ex-post (at the retailer level).

The AEMO rule change proposal does not describe the treatment of dispatchable LSUs in the FPP scheme, but it could be expected that these would have their deviations calculated individually[40], like scheduled generators. The demand deviation would then just cover non-LSU demand, and the FPP$ amounts would be calculated for non-LSU retailers exactly as described above.

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?

It has been assumed 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.

[39] implying the need for real-time metering of customers in the LSU

[40] which would imply the need for 4S metering on all customers in the LSU

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. AEMO will be able to advise on the materiality of these challenges.

Are all responses of demand to price regarded as demand response?

In this paper, demand response is defined very specifically as a response of demand (ie non-scheduled resources) to changes in the *spot* price. Demand can respond to other prices, of course. For example, some customers with rooftop PV and on conventional retail tariffs will buy a battery to soak up some of their PV output; allowing them to self-consume it later rather than exporting it to the grid. This is a rational response to the difference between the feed-in tariff and the consumption tariff. However, such a response is not considered to be demand response as the term is defined and used in this paper. The battery's operation depends on retail tariffs, *not* spot prices. Spot prices could rise or fall, and the battery operation would be completely unaffected.

On the other hand, if the battery was signed up to a VPP operated by the customer's retailer, say, its operation *could* then be affected by spot prices from time to time, via instructions from the retailer. This *would* be considered demand response.

There is a grey area where a retailer sets dynamic retail tariffs which may be indirectly influenced by the spot price. One example is a demand management tariff, where a retailer notifies a customer in advance of a "demand management period". If the customer reduces its demand over this period below its normal level, it receives a financial reward from the retailer.

The retailer would typically trigger a demand management period when its sees very high spot prices forecast in pre-dispatch. So the customer demand is responsive to spot prices, but to *forecast* spot prices rather than actuals. A retailer could make this type of response visible to AEMO by submitting a "vertical" quasi-bid. For example, if it predicts 100MW of response from demand management customers, it will submit 100MW of DR at all spot price levels, since the response will occur irrespective of whether the forecast spot prices eventuate.

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[41], but might place challenges on the retailer on whether and how to 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.

In examining this question, it is useful to distinguish between:

- *Retailer-initiated DR*: commonly referred to as virtual power plants (VPPs); and
- *Customer-initiated DR*: possibly supported by third-party "aggregators".

Retailers should reasonably be able to estimate the amount of DR provided by their VPPs. Indeed, this is largely the premise of VPPs. Admittedly, such estimation may be difficult in advance, even if it

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

can be calculated ex-post. But it seems likely that retailers can at least substantially improve on the current situation of the DR/VPP being invisible to AEMO.

Customer-initiated DR occurs through retailers passing through spot prices to their customers, who then either respond directly to these prices or engage an aggregator to do this for them. Retailers will have no direct information about this form of DR[42]. They would need to estimate or infer it, using static knowledge of these customers (eg size, type and location) to support ex-post analysis of customer actuals.

It would be a retailer's choice to offer a spot price pass-through product[43]. Retailers would only do so if they were confident in their ability to estimate the resulting DR[44]. This may lead to the emergence of specialist retailers who have the capability of estimating customer-initiated DR and the confidence to offer compelling spot price retail products.

Can retailers forecast their base demand too?

As discussed, the proposed design splits the role of demand estimation, with AEMO forecasting base demand and retailers 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[45]. Forecast errors will lead to deviations between target and actual generation which, in turn, determine the FPP$ amounts[46]. 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[47] 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[48].

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

---

[42] although they may offer load control algorithms for their customers to use, and so could have at least some indirect knowledge of response behaviour
[43] currently only a few, specialised retailers do this
[44] of course, a retailer could offer spot-price pass-through and simply pass-through the cost of invisible DR FPP$ to their customers. But retailer competition should lead customers to prefer retailers that are able to estimate and bid DR
[45] AEMO assesses the accuracy of these self-forecasts before permitting these to be fed into NEMDE
[46] and also the allocation of regulation FCAS costs
[47] ie each wind or solar farm
[48] and, as already noted, this is in fact based on the proxy of regional generation plus net imports

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 is notable that the dispatch mode in AEMO's proposed design implicitly requires self-forecasting for the LSU. However, it also envisages real-time metering for that LSU. As noted above, if this were to operate in parallel with the proposed design, AEMO would need to use this real-time metering to exclude the LSU from its regional demand forecasts.

Similarly, 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.

Can quasi-bids be safely input to NEMDE?

In the proposed design, quasi-bids are fed directly into NEMDE. This differs from AEMO's design, where analogous DR information provided under AEMO's visibility mode is *not* directly fed into NEMDE. This likely reflects AEMO's concern about the reliability and accuracy of these bids. If they are grossly inaccurate, this could adversely impact dispatch efficiency and even system security.

The issue of SSG self-forecasting was discussed above. Its introduction raised similar concerns: ie inaccurate SSG self-forecasts could adversely impact on dispatch. Reflecting this, AEMO was cautious in admitting these self-forecasts to NEMDE. It would receive these self-forecasts for a probationary period, long enough to be assured of their accuracy, before allowing them to be submitted to NEMDE.

A similar approach should be taken for quasi-bids. As discussed, the proposed design does not measure the accuracy of these bids directly, which is not possible, but rather calculates their impact on the accurate of the retailer-level demand forecasts. Accurate bids should lead to lower forecasting errors; grossly inaccurate bids would likely lead to *higher* errors.

AEMO could use this feature to assess the accuracy of bids over a probationary period, in which the bids are either not submitted to NEMDE or are scaled back to an amount which could not adversely impact on dispatch. For example, bid amounts for a retailer could be limited to no more than 50MW, say, over the probationary period.

Might retailers be incentivised to falsely bid DR in order to manipulate spot prices?

As noted, DR could materially impact spot price outcomes. Of course, this is to be welcomed if the quasi-bid accurately reflects DR, since this improves spot pricing efficiency and stability. But what if the bid is inaccurate? Worse still, what if a retailer deliberately manipulates its DR bids in order to achieve desired spot prices. For example, a retailer short of hedge cover might falsely bid DR to quell spot price spikes, even if it didn't actually have any price-responsive customers.

If it did this and the fake DR was cleared, then AEMO would then end up dispatching too little generation, frequency regulation would be called upon, and there would be a high demand deviation and associated FPP$. The retailer's false DR bid would lead to large errors in forecasts of its demand forecasts, meaning it would be charged a disproportionate share of the FPP$ that has already been inflated by its fake bids. This is the mechanism through which inaccurate bidding of DR – whether inadvertent or deliberate - will be discouraged.

But is this disincentive strong enough to discourage this false DR bidding? Or would the retailer still be financially ahead, despite the FPP$ penalty?

The FPP in its current form has been designed to encourage generators to conform with dispatch and mandatory PFR obligations, and even to voluntarily provide additional frequency regulation. However, compliance obligations still exist; the FPP complements rather than replaces them. The FPP incentives on retailers in the proposed design, do not have any analogous compliance backstop. Indeed, as noted, it is not possible to require "dispatch conformance" of DR in real-time, because the retailer has neither a dispatch target to follow nor real-time-metering to check that it does.

Instead, to address this concern over fake DR, quasi-bidding should be required to be in "good faith", analogous to generator bidding today. The AER would monitor the retailer demand forecasting errors calculated by AEMO to identify possible false or misleading quasi-bidding and take appropriate action.

What timescale should DR bids cover?

The objective of the proposed design is to reduce supply-demand imbalances – and so frequency regulation costs – caused by invisible DR. Strictly speaking, this only necessitates DR being bid immediately prior to dispatch. That would leave DR invisible to pre-dispatch (PD) and PASA, and the problem of poor spot pricing – or rather spot price *forecasting* - would remain in these timescales.

This could be addressed by requiring quasi-bidding into PASA or, at least, PD. Whilst quasi-bidding would remain voluntary, "good faith" rebidding rules could require that, if a retailer wanted to quasi-bid into dispatch, it would also have to quasi-bid into pre-dispatch. As for generators, leaving bidding to the last minute would not be permitted, unless this arose from a genuinely unforeseen change in circumstances.

There is a balance to be struck here. If PD quasi-bidding placed too much cost and inconvenience on retailers, they might decide not to bid at all, leaving DR invisible. Furthermore, DR bids might be intrinsically inaccurate day-ahead, say, and so requiring such bidding might not add much value.

Quasi-bidding into PASA exacerbates these concerns. An alternative would be for AEMO to estimate DR for PASA, based on analysis of dispatch quasi-bids. Since PASA is concerned with quantities rather than prices, AEMO "bidding" DR into PASA would likely be uncontentious.

Financial incentives for PD quasi-bidding could potentially be created by adapting existing RERT and RRO mechanisms. For example, bid DR might be exempt from the associated cost-recovery charges.

It is worth noting that the existing rules require retailers to provide static, longer-term, information on DR through AEMO's DSP portal. The proposed design needs to avoid duplicating this requirement.

Should visible DR 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 respectively – and the FPP$ charges allocated accordingly[49]. This section considers the arguments for and against its inclusion.

Chapter 6 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. A generator outage leads to a supply shortfall and a falling frequency. Whilst this fall is 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 DR 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;
- Correlate these deviations with frequency deviations; and

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

- Determine FCAS$ amounts based on these correlations.

However, a key difference, is that FCAS$ amounts are only attributed to those with positive correlations[50]: 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[51]. 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[52]. 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 not be used in the FPP$ calculations, and the customer/retailer gains no additional financial benefit from this metering.

[50] Which are *negative* correlations in AEMO's FPP algebra, because of the opposite sign convention already noted.
[51] In fact, it *has* to be non-linear in order to recover FCAS costs. Linearity would lead to it being zero-sum
[52] unless self-forecasting, as discussed above

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[53]. 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.

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

[53] 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

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.

Does it matter that frequency regulation generally covers multiple regions?

The proposed design is based around NEM regions. Retailers submit DR amounts for each region and AEMO carries out retailer demand forecasting by region. This reflects the fact that DR is a response to spot prices and spot prices are set regionally.

However, frequency regulation and FCAS is generally *not* regional. Frequency deviations reflect supply-demand imbalances across an AC island[54] rather than within an individual region. The FPP algebra, correspondingly, generally operates across AC islands. The demand deviation and associated FPP$ is calculated for each AC island, not each region.

In the proposed design, this means that the demand deviation will be decomposed at the AC island level and the respective FPP$ amounts will also apply to this level. Therefore, the allocation metrics will also be at this level:

- The wedge FPP$ is allocated in proportion to aggregate ΔDR across an AC island;
- The flat FPP$ is allocated in proportion to aggregate demand forecast error across an AC island; and
- The noise FPP$ is allocated in proportion to retailer demand across each an AC island.

Nevertheless, to minimise demand forecast errors, a retailer will need to bid DR at a regional level, because it doesn't know in advance what the regional spot prices are and so how much DR in each region will be cleared. Whilst a retailer might "get lucky" with opposite forecasting errors in different regions netting out when aggregated to the AC island, this does not fundamentally change the incentive to make DR visible through accurate, region-level bidding.

[54] Meaning regions interconnected by AC lines. The NEM has two AC islands - Tasmania and the mainland NEM - except under rare conditions where transmission outages mean that some mainland regions are islanded

Conclusions

The answers to the questions posed are summarised in table 12, below.

Question	Answer
Does the proposed design replace or complement AEMO's rule change proposal	It would enhance AEMO's visibility mode and could run in parallel with AEMO's dispatch mode
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?	No, only responses to spot prices; not responses to other prices such as retail tariffs.
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.
Can quasi-bids be safely input to NEMDE?	This will occur only once AEMO is confident that they are reliable and accurate.
Might retailers be incentivised to falsely bid DR in order to manipulate spot prices?	Possibly. Regulations should be introduced to prohibit this behaviour.
What timescale should DR bids cover?	In dispatch, to reduce frequency regulation; and possibly in pre-dispatch to improve scheduling.
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.
Does it matter that frequency regulation generally covers multiple regions?	No. The design can and will accommodate this in the settlement algebra.

Table 12: design questions and answers

8. 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 continuing 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 13, below.

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

Table 13: 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, if the proposed design were approved in the interim, the new FPP software could be designed to be ready to implement the alternative FPP design, with this extra functionality simply switched on when agreed.

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 decomposition would just be between the flat and noise components.

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 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.

9. 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.

AEMO is best place to advise on the nature and magnitude of these costs. However, 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. By aggregating and processing the quasi-bids into the same structure as scheduled load bids, there is no anticipated need to change NEMDE functionality. Calculated spot prices are applied to the original quasi-bids to determine cleared DR amounts for each retailer; the new "scheduled load dispatch targets" are used to determine cleared DR at the region level.

The retailer-level demand forecasting is a novel process, albeit based on existing 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 above, 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 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[55], 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.

[55] assuming quasi-bids are submitted to PD, as discussed in chapter 7

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. However, these costs are kept as low as possible by adapting and repurposing existing processes, rather than creating expensive new ones.

10. Overall Conclusions

To conclude:

1. Demand response (DR) is the response of non-scheduled resources to short-term changes in the spot price. 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 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, 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.