Australian Energy Market Commission

TECHNICAL REPORT

Optional Firm Access, Design and Testing

9 July 2015

REVIEW

Reference: EPR0039
Technical Report

Inquiries

Australian Energy Market Commission
PO Box A2449
Sydney South NSW 1235

E: aemc@aemc.gov.au
T: (02) 8296 7800
F: (02) 8296 7899

Reference: EPR0039

Citation

AEMC 2015, Optional Firm Access, Design and Testing, Technical Report, 28 May 2015, Sydney

About the AEMC

The AEMC reports to the Council of Australian Governments (COAG) through the COAG Energy Council. We have two functions. We make and amend the national electricity, gas and energy retail rules and conduct independent reviews for the COAG Energy Council.

Contents

1 Introduction

1.1 Purpose of this document

This document provides further detail on the proposed optional firm access (OFA) model, building on the description presented in Volume 2 of the Optional Firm Access, Design and Testing Review: Final Report.[1] However, this document can be read as a standalone description of the OFA model.

This document represents a complete, technical description of the optional firm access model as at 30 June 2015, which the Commission has developed, along with assistance from stakeholders.

The level of detail provided has been chosen with the objective of:

- presenting a complete picture of how the OFA model would operate;
- providing confidence that the model contains no irresolvable difficulties or inconsistencies;
- allowing stakeholders to analyse the potential impacts and implications for their organisations; and
- ensuring that OFA is detailed to a sufficient level in order to allow any further progression of the model.

1.2 Acknowledgement

The AEMC acknowledges the assistance of David Smith of Creative Energy Consulting in preparing this report.

1.3 Structure of this document

This document is structured as follows:

- Chapter 2 describes the fundamental concepts of access, firmness and optionality that provide the foundation and rationale for the model design, and introduces the model elements: the main building blocks of the model.
- Chapter 3 provides a top-down view of the model's scope and architecture, describing how the model elements interact with each other and with existing National Electricity Market (NEM) processes.
- Chapters 4 to 9 consider each of the model's main elements in turn.

1 AEMC, *Optional Firm Access, Design and Testing Review,* Final Report, 9 July 2015. All subsequent references to the Final Report mean this document.

- Appendices A through F provide further detail on each of the model's main elements.

Those chapters describing the model design (chapter 2 and chapters 4 to 9) are each divided into three subsections:

- The first subsection presents the *what*: a high-level description of the scope and functionality of the particular element.

- The second subsection presents the *how*: a blueprint of the element's design.

- The third subsection presents the *why*: design issues and options arising, and the rationale for selecting the proposed design.

2 Access

2.1 Overview

In the present NEM design, a generator is paid the regional market price on its dispatched output, irrespective of its location within a region. That is to say, its *access* to the regional market always equals its *dispatch level:* if the generator is dispatched, it automatically gets access; if it is not dispatched, it gets no access.

This linkage – between regional access and dispatch – is so intrinsic to the NEM design that it is easy to forget that it is a design *choice:* a choice that was made for good reasons during the original NEM development but that is neither inevitable nor irrevocable. Indeed, most electricity markets around the world do not link regional access to dispatch in this way.

This design choice creates operational and commercial issues, relating to congestion management and access certainty, which have been subject to numerous debate since the start of the NEM. The OFA model breaks the linkage and establishes a process for determining access *independently* from dispatch. Generators who require access certainty can procure a new *firm access service* from their local Transmission Network Service Provider (TNSP) and receive preferential access in return. The market's dispatch process is unchanged, with dispatch priority based on offer prices. Just as dispatch does not affect access, access does not affect dispatch.

A generator *without* access receives a *local* price for its output. A generator *with* access receives the regional price for its output; it is compensated – based on the difference between the regional and local prices – to the extent that its output is below its access level.

An analogous situation exists in relation to interconnectors. When interconnectors are dispatched – to flow power between two neighbouring regions – revenue is paid to the interconnector based on the value of transferring power from the lower-price to the higher-price region. In a sense, interconnectors then have access to the *inter-regional* market in a similar way to a generator gaining access to the regional market. But, as with generators, access equals dispatch: if the interconnector is not dispatched it earns no revenue. OFA again breaks this link, allowing interconnectors to earn revenue even when they are not dispatched. This is achieved through compensating an interconnector that is dispatched below its access level.

Access – like dispatch – is constrained in *aggregate* by the size and reliability of the transmission network. TNSPs are therefore required to plan and operate their networks to a new *firm access standard* which ensures that a guaranteed level of access firmness can be provided to those *firm generators* and *firm interconnectors* that have procured firm access.

The costs that this obligation creates for TNSPs are recovered from firm generators through *access charges*. These charges provide new locational signals for new generation: in choosing its location and firm access level, a generator will tailor its access cost and firmness to its budget and risk appetite.

Interconnectors are different to generators in that they have no obvious owner to whom access-related revenues should be paid. In the current arrangements, market participants can purchase, through the Settlement Residue Auction, the right to receive interconnector revenues. The OFA model introduces Firm Interconnector Rights which are analogous to these existing rights (but firmer). Any market participant can buy firm interconnector rights so long as they pay the associated access charges.

2.2 Design Blueprint

2.2.1 Generation Dispatch and Pricing

NEM dispatch is conventionally thought of as a regional market clearing process operating as follows:[2]

1. Generators submit dispatch offers to the Australian Energy Market Operator (AEMO) which represent the lowest price at which they are willing to be dispatched.

2. The NEM dispatch engine (NEMDE) determines the price at which sufficient generation can be dispatched so as to meet regional demand.

3. That price is the regional price, (ie, the regional reference price or RRP), which is paid to all dispatched generators.[3]

The above is a reasonable description of market clearing when there are no transmission constraints interfering with dispatch: or, conversely, where the dispatch determined through the process above does not overload the transmission network.

However, this is a poor description of NEM dispatch in the situation where transmission constraints become relevant. In this case, a more accurate description would be:

1. Generators submit dispatch offers to AEMO, which NEMDE *interprets* to be the lowest *local* price at which they are willing to be dispatched.

2. NEMDE determines a local price at each node such that:

    (a) sufficient generation is dispatched to meet regional demand;

    (b) the transmission network is not overloaded; and

---

2 It can also be thought of as merit-order dispatch, with the regional price set at the offer price of the marginal generator.

3 Transmission losses are ignored in this discussion and in general in this document. They are not pertinent to the OFA model, which does not change the way they are calculated and applied.

(c) subject to the above two conditions, total dispatch costs (as represented in dispatch offers) are minimised.

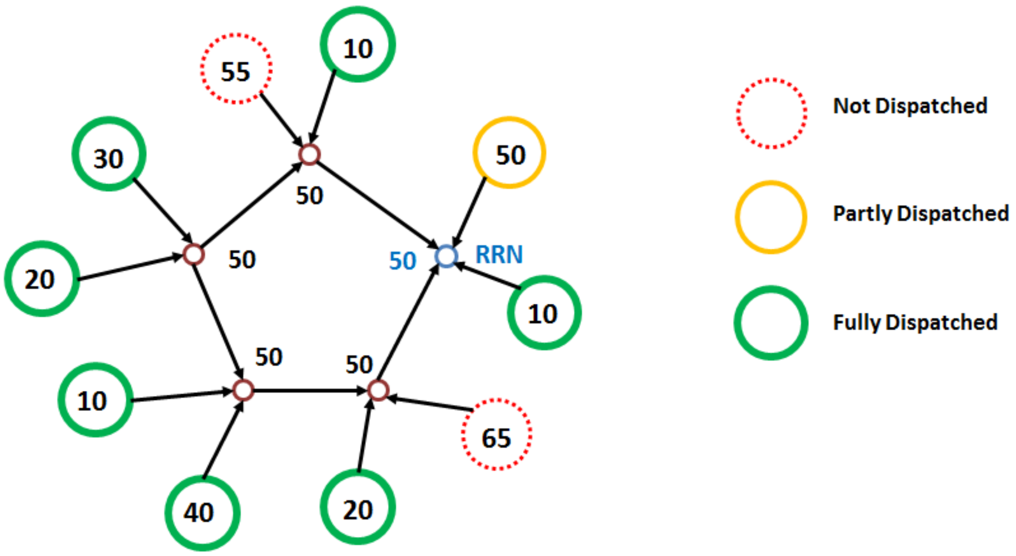
3. The regional price (RRP) is *defined* to be the local price at the regional reference node (RRN).

4. The RRP is paid to all dispatched generators.

The local price calculated through this dispatch process is referred to as the locational marginal price (LMP).

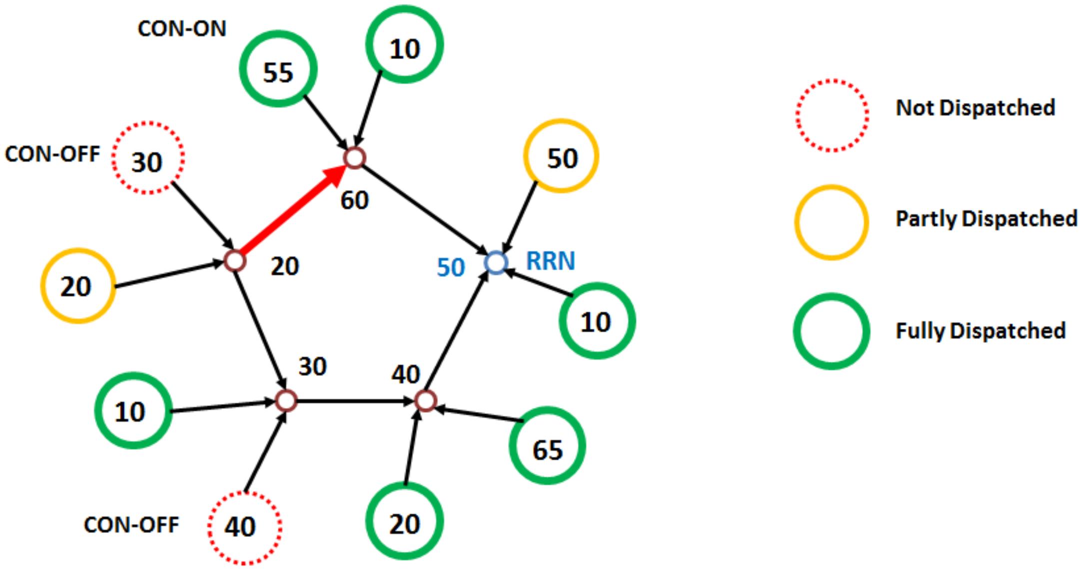
Figure 2.1 and Figure 2.2 illustrate these concepts. In each figure there are several generators connected at points around a loop in the transmission network. The numbers inside the circles represent generator offer prices. The numbers inside the loop represent the local prices. The RRN is marked and the RRP is the local price at that node.

Figure 2.1 Uncongested dispatch



In the uncongested dispatch, the local prices are all the same. So although each generator is dispatched according to its local price, this is indistinguishable from each generator being dispatched according to the RRP: the local price at the RRN.

Figure 2.2 Congested dispatch



In Figure 2.2 there is congestion on one transmission line (highlighted in red). This causes the local prices to diverge. Each generator is still dispatched according to its local clearing price. But, when dispatch is assessed against the RRP, there are some generators that are:

- *constrained on*: dispatched despite their offer price being higher than the RRP; or
- *constrained off*: not dispatched despite their offer price being lower than the RRP.

What causes these generators to be constrained on or off is the difference between local prices and regional prices. In the NEM there are differences between:

- *NEM dispatch*: which is a local market clearing process; and
- *NEM settlement*: which is designed to reflect a regional market clearing process.

2.2.2 Access Definitions

Access within a Region

A framework for resolving this inconsistency is to consider that a generator's access to the NEM is made up of two components:

1. *Dispatch access*: which gives a generator a right to submit a dispatch offer, be dispatched at its local node in accordance with that offer[4] and be paid the local price for its output.

4 Ie, dispatched if the local price exceeds its offer.

2. *Network access*: which gives the generator the right, notionally, to buy an amount of power at its local node at the local price, transport it over the transmission network and sell it at the regional price.

In this model, the settlement payment to a generator is:

$$Pay\$ = Pay\$_{dispatch} + Pay\$_{network} = LMP \times G + (RRP - LMP) \times A \qquad (2.1)$$

Where:

G = dispatched output

A = amount of network access

RRP = regional price (regional reference price)

LMP = local price (locational marginal price)

In the current NEM design, the level of network access provided is always set equal to the dispatch level (A=G) and so equation (2.1) simplifies to become the more familiar:

$$Pay\$ = RRP \times G$$

Dispatch access is a *physical* service: a generator must be connected to the transmission network so that it can be dispatched, and the generator must actually run to receive payment. On the other hand, network access is a *financial* service: from the generator's perspective, it is simply an additional payment from AEMO, not (explicitly) relying on transmission or generation. These physical and financial aspects of access can, in principle, be separated. There is no fundamental reason why the access level needs to equal the dispatch level.

In summary, it is possible to change arrangements for the provision of network access without making any corresponding changes to dispatch access: ie, to the dispatch process. The access level, A, can be changed without impacting the dispatch level, G.

Since it is *network* access that is the focus of the OFA model, this is generally referred to simply as *access* in the remainder of this document.

No-regret Dispatch

A generator will have some variable costs associated with being dispatched. Assume for simplicity that these are linear with output:[5]

$$Variable\ cost\$ = C \times G \qquad (2.2)$$

Where:

C = short-run marginal cost of generation (or SRMC)

5 Although the conclusions of this section are unchanged if this assumption is relaxed.

The operating profit[6] of a generator can be defined as the difference between its NEM revenue and its variable costs. Combining equations 2.1 and 2.2 above gives:

$$\text{Operating Profit\$} = (\text{LMP-C}) \times \text{G} + (\text{RRP-LMP}) \times \text{A} \qquad (2.3)$$

If the access level, A, is independent of dispatch, then the first term in equation 2.3 represents the extra profit arising from dispatch. If a generator bids at or above its cost, C, then it will only be dispatched when LMP≥C and so dispatch can only *increase* its profitability. This situation is referred to as *no-regret* dispatch and is an important principle of the OFA model. The following conditions are, together, sufficient to ensure no-regret dispatch:

- the access level is set independently of the dispatch level;
- the local price used in settlements is a *clearing price*: ie, the generator is only dispatched if the local price is equal to or higher than its offer price; and
- the generator's offer price is no lower than its generation SRMC.

Access between regions

The previous section discusses access for generators within a region: access from a generator's local node to the RRN in the *same* region. So, a generator connected in NSW, say, has (network) access to the NSW RRN (in Sydney). This access can be referred to as intra-regional, since it applies *within* a single region.

The NEM comprises multiple, interconnected regions. Notional entities, referred to as *interconnectors*, flow power between regions, from a RRN in one region to the RRN in a neighbouring region.

Equation 2.1 can be adapted to the inter-regional situation as follows:

G=0 since an interconnector does not itself generate power

A_{IR}=amount of (inter-regional) network access that the interconnector enjoys

RRP_M = regional price in the importing region: the region of the RRN *to* which the interconnector has access

RRP_X = regional price in the exporting region: the region of the RRN *from* which the interconnector has access

Making these changes, we obtain an inter-regional equation that defines the settlement payment to an interconnector:

$$\text{Pay\$}_{IC} = (RRP_M - RRP_X) \times A_{IR} \qquad (2.4)$$

6 In accounting terminology, this is commonly referred to as "earnings before interest, taxes, depreciation and amortization" or EBITDA.

As before, losses are ignored for simplicity.

In the current arrangements, the access amount is defined to be equal to the dispatched interconnector flow, *I*, giving the familiar equation for the inter-regional settlement residue (IRSR) payment to the interconnector:

$$Pay\$_{IC} = (RRP_M - RRP_X) \times I$$

Interconnectors can flow power in either direction, and so it is not clear which region should be importing and which exporting. To clarify this, an entity called a directed interconnector (DIC) is introduced. Each interconnector (eg QNI)[7] has a pair of associated DICs (ie, "QNI north" and "QNI south"), oppositely directed. With this direction attribute, the exporting region and importing region are clearly defined.

Table 2.1 Directed Interconnectors in the NEM

DIC	Exporting Region	Importing Region	Type
QNI North	NSW	QLD	AC
QNI South	QLD	NSW	AC
Terranora North	NSW	QLD	DC
Terranora South	QLD	NSW	DC
Vic-NSW	VIC	NSW	AC
NSW-Vic	NSW	VIC	AC
Heywood West	VIC	SA	AC
Heywood East	SA	VIC	AC
Murraylink West	VIC	SA	DC
Murraylink East	SA	VIC	DC

Note that Basslink is *not* included in this table because it is a Market Network Service Provider (MNSP), a merchant interconnector, rather than a *regulated* interconnector. OFA treats MNSPs as generators in the importing region, rather than as interconnectors, and so MNSPs experience intra-regional rather than inter-regional access.

There are some *parallel interconnectors* in the table: ie, interconnectors flowing between the same two regions. A DC interconnector[8] is dispatched independently of the parallel AC interconnector and so is treated separately under OFA. However, parallel

7 The Queensland-NSW Interconnector.

8 An interconnector associated with a regulated DC transmission line that crosses a regional boundary.

AC interconnections cannot be independently dispatched and so these are regarded as a single, composite interconnector.[9]

It is seen from equation 2.4 above that payments will be made in relation to an interconnector whenever the regional prices at the ends of the interconnector diverge. It then needs to be decided to which of the two associated DICs the payment is made. In general, the payment is to the DIC that is directed towards the higher price region: such that $RRP_M > RRP_X$. However, there are exceptions to this rule, which are discussed further in section 4.2.2.

2.2.3 Access Settlement

Intra-regional Access Settlement

Equation (2.1) above can be rewritten as follows:

$$Pay\$ = RRP \times G + (RRP - LMP) \times (A - G) \qquad (2.5)$$

In this formulation, the first term is exactly the same as the generator settlement payment in the existing NEM design.[10] The second term represents the additional payment that must be made to separate access from dispatch, as discussed above.

This new payment is calculated in a new process that is introduced under OFA, referred to as *access settlement.* This process is undertaken by AEMO, alongside the existing *regional settlement* that calculates the first term. The two payments are aggregated and settled together, using existing settlement processes.

A fundamental principle of the OFA model is that *aggregate* settlement payments must *balance* in each settlement period:[11] this principle is discussed in section 2.3.4. Since regional settlement balances,[12] the new access settlement must also balance.

The implications of this principle are discussed below for a simple two-node network, shown in Figure 2.3. The more general situation of a meshed, multi-node network is discussed in section 4.3.1.

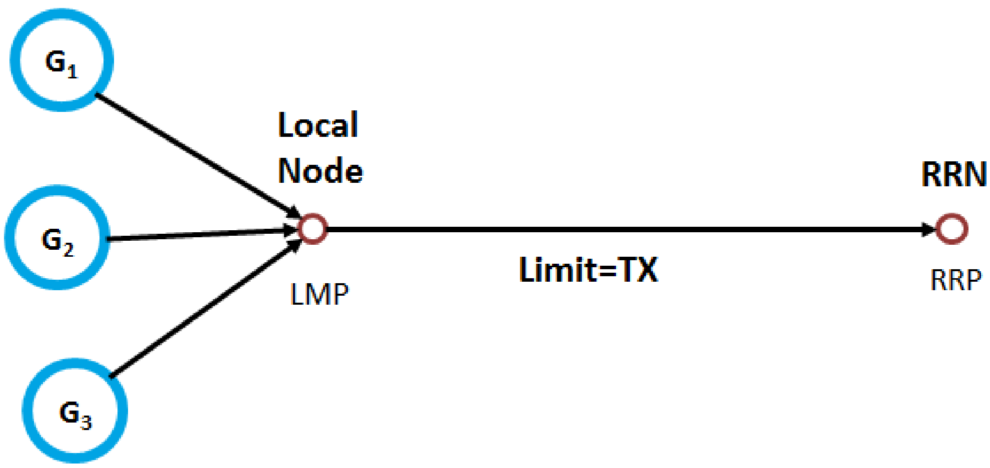
9 For example, the NSW-Vic interconnector incorporates three separate crossings of the NSW-Victoria boundary.

10 Again, ignoring transmission losses for simplicity.

11 The settlement period is a 30-minute trading interval, discussed further in section 4.3.7.

12 Including the paying out of the inter-regional settlement residue and ignoring losses.

Figure 2.3 Two-node network example



In this simple example, all generators are connected to the same node, so LMP is the same for each generator.[13] Dispatch and access will vary between generators, so define:

A_i = access for generator i

G_i = dispatch for generator i

Then, from equation (2.2), the access settlement payment to generator i is:

$Pay\$_i = (RRP - LMP) \times (A_i - G_i)$ (2.6)

The total payment across all generators is then:

$Total\ Pay\$ = (RRP - LMP) \times (A_T - G_T)$

Where:

A_T is total access

G_T is total generation

If the transmission line connecting the two nodes is uncongested then RRP=LMP and so the access settlement payments are zero: individually and collectively. On the other hand, if the line is congested and the prices diverge then, for settlement to balance, we must have:

$A_T = G_T$

But, since the line is congested we know that:

13 There must also be a generator connected to the RRN that sets the RRP but this generator does not participate in access settlement and so is ignored in the analysis below.

$$G_T = TX$$

Where:

TX = transmission capacity

Therefore, a sufficient condition for settlement balancing is:

$$A_T = TX \tag{2.7}$$

Thus a critical element of the setting of generator access levels involves sharing, or *allocating*, the available transmission capacity between various generators. A similar, generalised result holds on a meshed network with multiple points of congestion, as explained in the section 4.2.1.

Inter-regional Access Settlement

To obtain a level of inter-regional access, A_{IR}, a DIC must be paid:

$$DIP\$ = A_{IR} \times (RRP_M - RRP_X)$$

Where:

$DIP\$$ = payment to directed interconnector

This equation can be rewritten as:

$$DIP\$ = I \times (RRP_M - RRP_X) + (A_{IR} - I) \times (RRP_M - RRP_X)$$

Where:

I is the flow on the interconnector

The first term is equal to the existing IRSR payment.[14] Therefore, this payment must be supplemented in access settlement by an amount equal to the second term:

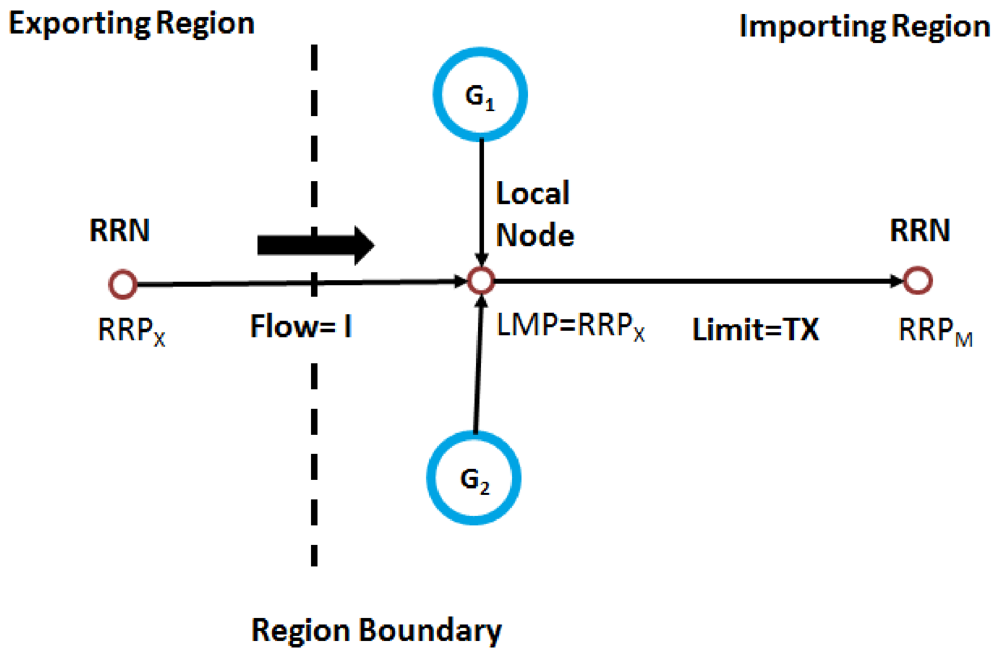
$$Pay\$_{IC} = (A_{IR} - I) \times (RRP_M - RRP_X)$$

Figure 2.4 extends the two-node example to a three-node example involving two RRNs and a DIC.

---

14 Ignoring losses, as usual.

Figure 2.4 Inter-regional network example



For simplicity, it is assumed that there is no congestion between the exporting RRN and the local node, so the local prices are the same at these two nodes: RRP_X = LMP. Substituting for LMP in the usual intra-regional access settlement formula, the payment to each generator i from access settlement is:

$$Pay\$_i = (RRP_M - RRP_X) \times (A_i - G_i)$$

Therefore the total access settlement payment is:

$$Pay\$ = \sum_i Pay\$_i + Pay\$_{IC} = (RRP_M - RRP_X) \times (\sum_i A_i + A_{IR} - \sum_i G_i - I)$$

Therefore, for access settlement to balance (Pay$=0) we require that:

$$A_T = G + I$$

Where A_T is total intra-regional *and* inter-regional access.

$$A = \sum_i A_i + A_{IR}$$

When the line is uncongested, the two RRPs are the same and so all access settlement payments are zero. When the line is congested:

$$G+I = TX$$

And so, a sufficient condition for access settlement to balance is:

$$A_T = TX$$

Thus, the fundamental constraint on aggregate access is the same for intra-regional and inter-regional constraints: total access must equal transmission capacity.

2.2.4 Firm Access

Overview

The OFA model removes the existing link between network access and dispatch access. For each generator, the dispatch level is determined as it is currently.[15] The level of network access is set *independently* of dispatch. Which raises the question: what should the access level be based on instead?

As indicated by its name, the OFA model introduces a concept of *firm access*, a service that TNSPs provide to generators and DICs. Within the overall limitation that total access must equal transmission capacity, priority is given to allocating access to those *firm generators* and firm DICs (referred to generically as *firm participants*) to whom firm access is being provided.

Registered Access

TNSPs issue intra-regional firm access to generators through various *issuance* processes described later in this section. AEMO is notified of the MW amounts of firm access issued to each generator and AEMO then records these amounts in a *firm access register* that it maintains. These registered access amounts are used by AEMO in allocating access for access settlement.

However, TNSPs do not, correspondingly, issue inter-regional firm access to DICs, since DICs are notional settlement entities rather than market participants. Instead, TNSPs issue *firm interconnector rights*, or FIRs, to market participants (ie, generators, retailers or MNSPs). FIRs give the holder the right to receive a corresponding share of the settlement amounts paid to the associated DICs.[16] The registered access for a DIC is the aggregate volume of all FIRs issued in relation to that DIC. FIR holdings and registered inter-regional access are also recorded by AEMO on the firm access register.

Although access settlement amounts are calculated every thirty minutes, it is unlikely that registered access would vary this frequently. When firm access is issued it will have a defined term, with the start and end of the term will be recorded in the firm access register.

Each issuance of firm access is defined by a set of service parameters, which AEMO records in the firm access register. These parameters are listed in Table 2.2 below.

---

15 There is no change to the dispatch process. However, changes to bidding incentives will lead to changes in dispatch outcomes.

16 In this respect they are similar to the Settlement Residue Auction (SRA) rights issued under current arrangements.

Table 2.2 Service Parameters of Registered Access

Service Parameter	Intra-regional Firm access	Inter-regional Firm access
Holder	Generating Company to whom the firm access is provided	Market participant who holds the FIR
Location	Generating unit to which access is provided	DIC on which the FIR is held
Amount	MW volume of firm access	MW volume of FIR
Term	Start and end of firm access term	Start and end of firm access term

Access Allocation

Network access is allocated according to two principles:

1. priority is given to firm participants; and

2. total access allocated must equal transmission capacity.

In the event that the total registered access is less than transmission capacity, the remaining transmission capacity is allocated to non-firm generators and non-firm DICs: ie, those who do not have any registered access.

On the other hand, if total registered access is *greater* than transmission capacity, then access is scaled back: each firm generator and firm DIC only receives a *proportion* of its registered access amount. In this situation, non-firm generators and non-firm DICs receive no access.

This is a simplified description. A fuller definition of the access allocation process is provided in section 4.2.4.

Although registered access will only change occasionally (ie, when some existing firm access expires or is sold, or some new firm access commences), transmission capacity can vary continually. Therefore, within access settlements, network access is reallocated for each settlement period.

Access Firmness

A firm generator or DIC may be allocated its full, registered access amount in some settlement intervals but have its access scaled back in others. The frequency and severity of this scaling back will depend upon the level of transmission capacity relative to the aggregate of registered access.

The level of access allocated relative to generator capacity[17] is referred to as the *access firmness*. Clearly, a firm generator will enjoy a much higher level of access firmness than a non-firm generator, due to its prioritisation in the access allocation.

Firm Access Standard

Equation (2.7) means that, overall, access can only be as firm as the transmission network: at the extreme, if there is no transmission capacity, there can be no access. In the OFA design, a *firm access standard* is established which requires each TNSP to make sufficient transmission capacity available to provide a specified level access firmness to all firm generators and firm DICs, individually and concurrently. This requirement is called the firm access standard (FAS). The FAS requires TNSPs plan to provide firm generators and DICs with at least a specified level of access firmness. The FAS takes no account of non-firm generators and non-firm DICs.

Even for firm generators and DICs, access is required to be firm but not fixed. That is to say, some scaling back of firm access is permitted and expected. The FAS is discussed further in chapter 5.

Access Charges

Because a TNSP is required to expand and maintain its transmission network in accordance with the FAS requirement, it incurs costs in providing firm access service to generators and DICs.[18] This cost is recovered from firm generators and from FIR purchasers, through an access charge. The access charge is determined when new firm access is agreed. It is fixed[19] for the life of the firm access agreement, to ensure maximum financial certainty for firm generators and FIR holders. The charge is based on the forecast incremental cost associated with the new access.

A TNSP has no FAS (or other) obligations in relation to *non-firm* generators and so their presence does not directly create any additional costs. Therefore, non-firm generators do not pay access charges.

Access charges are discussed further in chapter 6.

Firm Access Issuance

There are three categories of processes associated with access issuance:

- long-term;
- short-term; and

17 Allocated access never exceeds generator capacity, as discussed in section 4.2.4.

18 Unless there is so much spare capacity, the new access does not affect future transmission expansion.

19 Except for some defined indexation.

- transitional.

Under a *long-term* process, a TNSP may need to expand its transmission capacity to cover the additional firm access. The recipient(s) of the firm access are therefore required to pay access charges to cover the associated TNSP costs, as described in the previous section.

Under a *short-term* process, there is no possibility of transmission expansion, because the term of the issued firm access does not give a TNSP sufficient time to arrange for such expansion. This means that there is no cost imposed on the TNSP. It also means that the amount of short-term issuance is limited to existing spare transmission capacity: ie, any existing capacity over and above that required to meet the FAS. An auction process, with no reserve price, is used in order to ration this limited amount efficiently between market participants.

The *transitional issuance* process occurs only at OFA commencement and is used to allocate between existing[20] generators the firm access that can be provided by the existing network.

The transitional issuance process is described in chapter 9 and the other issuance processes are described in chapter 7.

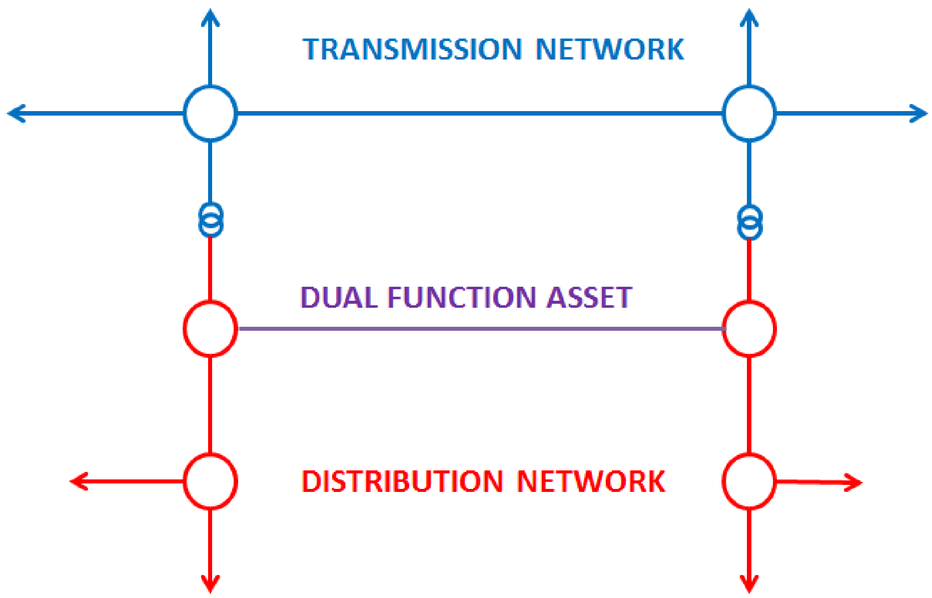
2.2.5 Scope of OFA

Network Scope

OFA is designed to apply only to the shared transmission network. However, in some cases, a distribution line can run in parallel with a transmission line, meaning that a limitation on one can limit flows on the other. Such distribution lines are referred to in the Rules as *dual function assets* and such assets are included in OFA in the same way as if they were transmission assets. Other distribution assets are not covered by the OFA model. An example of a dual function asset is presented in Figure 2.5.

20 At OFA commencement.

Figure 2.5 Scope of OFA



The FAS obligations only apply to the TNSP in a region that has planning obligations. That TNSP may need to enter into agreements, or coordinate its network activities, with other network operators in the region, in order to meet those obligations.

Market Scope

In the NEM, generators may be categorised as *scheduled*, *semi-scheduled* or *non-scheduled*. The first two categories must comply with dispatch instructions issued by AEMO; non-scheduled generators choose their dispatch level themselves. The OFA model covers only scheduled and semi-scheduled generators. Non-scheduled generators do not participate in access settlements and do not receive a firm access service. This means, in effect, that the access level they receive always equals their dispatch level and so they are always paid RRP.

Generators are also categorised as either *market* or *non-market*. The former participate in AEMO settlement whereas the latter are paid by their local retailer[21] who, in turn, settles with AEMO. As is the case currently with regional settlement, the local retailer for a *non*-market scheduled (or semi-scheduled) generator will participate on the generator's behalf.[22]

As discussed in section 2.2.3, regulated interconnectors participate in access settlement, as a pair of DICs, and their payments are shared between FIR holders.

21 It is also permitted for a non-market generator to sell directly to an end-customer at the same connection point.

22 Although this approach might not be effective where the non-market-scheduled generator sells directly to an end-customer.

Merchant (MNSP) interconnectors participate in the importing region (based on the direction of the interconnector flow in a settlement interval) in an analogous way to scheduled generators. They do not participate in the exporting region.

Customers (scheduled and non-scheduled) do not participate in the OFA model.

2.3 Design Issues and Options

2.3.1 Overview

In its specifics, the OFA model is unlike any electricity market design that has been implemented. One will not find terms such as “firm access” and “flowgate support” in the electricity market design literature. But at a fundamental level, the model has strong similarities with a generic electricity market design, referred to in the US as the Standard Market Design (SMD), variations of which have been implemented across North America.

There are two principal design elements that OFA shares with the SMD:

- generators are paid for their dispatch output at local prices based on the LMPs that are calculated in the dispatch process; and
- financial instruments are issued which allow generators to hedge the difference between their local price and the (wholesale market) price paid by customers: these are referred to in the SMD as Fixed Transmission Rights (FTRs).

It is helpful to compare the design elements of the OFA model with those of the SMD. Obviously, given that the OFA has been designed and customised for NEM implementation, whilst the SMD reflects its American context, design differences are to be expected. Examining those differences, and assessing whether and how they reflect the contextual differences, can help to explain how the key OFA design decisions were made: why some design options were adopted and others were rejected.

This is not to say that the OFA has been deliberately modelled on the SMD. Indeed, in its initial concept and early manifestations (in the Transmission Frameworks Review (TFR)) it was quite different to the SMD. The final OFA design, presented here, is closer to the SMD, but substantial differences remain, which are discussed in this section.

The design elements of OFA and SMD considered in this section are:

1. access settlement;
2. regional pricing;
3. settlement balancing; and
4. transmission planning and reliability.

These are discussed in turn, below.

A further important difference between OFA and SMD is the treatment of flowgate support. This is discussed in chapter 4.3.3.

2.3.2 Access settlement

Overview

Two formulations of generator payments under OFA were presented previously:

$$\text{Pay\$} = \text{Region pay\$} + \text{access pay\$} = \text{RRP} \times \text{G} + (\text{RRP} - \text{LMP}) \times (\text{A} - \text{G})$$

And:

$$\text{Pay\$} = \text{Dispatch Access\$} + \text{Network Access\$} = \text{LMP} \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A}$$

For reasons discussed below, the first formulation will be referred to as the *compensation formulation* and the second the *SMD formulation*. Although the two formulations are mathematically identical, they lead towards two different interpretations of OFA. They are discussed in turn below.

Compensation Formulation

In this formulation, generators continue to be paid the RRP for their output. Intra-regional constraints can lead to a generator being constrained off: ie, not dispatched, despite its offer price being below the RRP.

A constrained off generator loses access to the RRP. The OFA model has its origins in clause 5.4A of the Rules, which aimed to provide a mechanism to compensate generators for being constrained off. The second term in the compensation formulation would seem to provide this compensation.

There is a problem with this interpretation, however. A generator being dispatched below its access level during congestion – and so receiving a payment from the second term - does not necessarily mean it is constrained off. It might, instead, be unavailable or out-of-merit.[23] Compensation is provided, but it is poorly targeted.

The difficulty is that, in order to determine whether a generator is genuinely constrained off and, if so, what level of compensation is appropriate, a generator's actual costs need to be known or estimated. But in a bid-based market, in which generators are not obliged to reveal their costs, such cost estimation is problematic. Relying on the offer price as a proxy for cost will not work, because an out-of-merit generator can rebid a lower offer price so that it looks to be constrained-off.[24] This issue is discussed further in section 4.3.5.

23 Ie, have an offer price higher than RRP.

24 It would do this by bidding an offer price between its LMP and the RRP. This ensures that it is not dispatched.

In the current NEM design, generators receive cost-based compensation when they are directed by AEMO. But, in this case, an independent expert is engaged to estimate the associated cost impact. This is possible for directions, because they only occur rarely, but would not be practical for constrained-off generation, which occurs frequently.

A second concern with the compensation approach is that, since each generator has different costs, the level of compensation payable – and so the value obtained from received firm access – would vary, even between generators located at the same node. A mid-merit generator might be compensated frequently; a peaking generator rarely.

It would be unreasonable to charge the two generators the same amount for firm access, but it is unclear how to price access in a way which reflects the cost differences.

In the light of these difficulties, the final OFA design pays "compensation" similar to what an FTR holder would receive in the SMD. This is discussed next.

SMD Formulation

In this formulation, the generator payment is:

$$\text{Pay\$} = \text{LMP} \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A}$$

Generation is *nodally* priced: ie, there is a different price at each generator node, based on the LMP. RRP is the price paid by customers: this is discussed further in section 2.3.3. Each generator has an amount, A, of FTRs, which provide a hedge between its local price and the customer price.

In this formulation, the OFA can appear to be quite a radical change to the existing NEM design, changing it from a regionally-priced to a nodally-priced market. Radical, perhaps, from a NEM perspective. But internationally, nodal pricing is the common design and regional pricing – of the form used in the NEM – is unknown.

In a nodal market, there is no concept of generators being constrained-off. A generator is always dispatched against the local price in accordance with its offer: ie, dispatched when its offer price is at or below the local price. So there is no question of compensation. Firm access simply becomes a financial instrument that hedges differences between the regional and local prices.

Conclusions

Although access settlement is commonly explained in terms of compensation payments, this reflects OFA's origins rather than its final design. Settlement payments under OFA are now the same as those applying under the SMD. This convergence between the two designs stems from the fact that the design benefit of basing generator payments on LMP is independent of context and arises in the NEM as well as in North America.[25]

2.3.3 Regions

Demand-side Pricing

Under SMD, customer pricing may be nodal or zonal. In the former, customers pay the local price (LMP) at their local node. In the latter, zones are defined geographically, somewhat analogously to NEM regions. Customers pay the zonal price, which can be defined as:

- the average of the nodal prices in the zone; or
- the LMP calculated under a *notional dispatch* in which all intra-zonal transmission constraints are removed.

The NEM uses a third approach to regional pricing: setting the regional price to be the LMP at the designated RRN. This approach is not used in any other electricity markets.

Intra-regional Access

In the OFA, intra-regional access means financial access from the local node to the RRN. This is equivalent to holding an FTR from the local node to the RRN. Under a zonally-priced SMD, generators will seek to have financial access (ie, FTRs) from their local node to the zonal price. There is no "zonal node" in an equivalent sense to the RRN, but there is some financial equivalence.

It can seem counter-intuitive, under the OFA model, that all generators seek access to the RRN since, physically, demand is spread across the region rather than residing just at the RRN. Defining access in this way creates a number of difficulties in the OFA design, which are discussed in the following chapters. But this design is a consequence of the design of regional pricing in the NEM. Adopting demand-side pricing more aligned with the SMD approaches would likely simplify the OFA design. But that would be a radical change to the NEM design that goes beyond the remit of OFA.

Regional TNSPs

Regional boundaries in the NEM align with the boundaries of the state-based TNSPs. This has not always been the case. There is no necessity for it either in the current NEM design or in the OFA, but it does allow for some simplicity.

In SMD markets, many or all of the TNSP functions relevant to OFA (specifically FTR issuance, planning and access pricing) are undertaken by a Regional Transmission

25 At least when LMP is less than RRP. The situation when LMP is greater than RRP is discussed further in section 4.3.3.

Operator (RTO) covering the entire market and having a scope of responsibilities similar to AEMO in Victoria.

This structural difference is most clearly reflected in the governance of access pricing. Under OFA, this role is given to the AER, as an independent, NEM-wide institution. Under SMD, this role would typically be given to the RTO, for similar reasons.

Inter-regional Access

Access from a node to another node under OFA is similar to having an FTR between those two nodes. Thus, inter-regional access is similar to holding an FTR between two RRNs. In zonal SMD markets, inter-zonal FTRs could similarly be issued. Alternatively, an FTR might be issued that provides access from a node in one zone to the zonal price in another zone. The OFA does not allow this: there is likely to be limited demand for the product, which would create substantial additional complexity in the OFA design.

Under OFA, region boundaries do not disappear but they do become less significant. Although region boundaries are transparent to the dispatch process[26], generator bidding strategies are driven by regional pricing and so are strongly dependent upon regional boundaries. OFA lessens, but does not remove, the influence of regions and helps to provide more of a level playing field between inter-regional and intra-regional trading.

Conclusions

The NEM is unique, in market design terms, in having regional reference nodes as the basis for regional pricing. In most regions, this approach reflects Australian geography in which most demand is located in the metropolitan region around each state's capital. So a conventional zonal price found in an SMD market might not look materially different to the RRP. Queensland and Tasmania are exceptions, given that they have substantial load remote from the RRN.

Some idiosyncrasies in the OFA design stem from this regional pricing approach.

2.3.4 Settlement Balancing

Overview

A principle of OFA is that settlement must balance within each settlement period (ie, half-hour), as discussed in section 2.2.3. SMD does not follow this principle and allows settlement surplus and deficits. The implications of these design decisions are discussed below.

26 Since inter-regional constraints are fully "co-optimised".

Fixed Access Approach

Under SMD, access is fixed:[27] payments on FTRs are always made on the same *access volume*. If, on a congested flowgate, total access exceeds flowgate capacity then a settlement deficit will arise.[28] On the other hand, if total access is below flowgate capacity, then a settlement surplus will arise.

These surpluses and deficits are managed by "smearing" them across settlement periods. Surpluses are paid into an accumulating pool of funds and deficits are funded by this pool. Where aggregate surpluses over time are insufficient to fund aggregate deficits, some additional funding is required, typically sourced from an uplift charge placed on customer prices. On the other hand, a long-term surplus would be returned to TNSPs or customers. So, settlement balances in the long-run, but not on a half-hourly basis.

Firm Access Approach

Under OFA, deficits are avoided by scaling back access when there is a flowgate shortfall. Access is now firm rather than fixed. Surpluses are avoided by allocating excess flowgate capacity to non-firm generators.

Choice of firm approach for OFA

There are several reasons why the firm access, rather than fixed access, approach has been adopted for OFA.

First, the NEM design has generally avoided allowing surpluses to accrue or requiring deficit-funding uplift charges on customers.

Second, the NEM transmission network is rather "stringy" meaning that transmission outages can create severe congestion and large flowgate shortfalls. The North American networks on which SMD operate tend to be strong and highly-meshed, meaning that transmission outages – when they occur – do not have the same impact.

Third, the NEM is an energy-only market meaning that extreme spot prices are needed from time to time to fund peaking plants. The SMD includes a capacity market which performs this role, meaning that extreme prices are not needed: typically the energy price cap in SMD markets will be substantially lower than the NEM market price cap.

These second and third factors mean that the relative magnitude of half-hourly settlement deficits and surpluses, under fixed access, would likely be much higher in the NEM than in SMD markets. This could lead to substantial deficits accruing over time and create the need for uplift charges on customers.

---

27 Hence the name of *fixed* transmission rights.

28 This can be seen from the algebra presented in section 2.2.3.

Conclusion

Settlement balancing is an important and necessary design principle in the NEM, due to its high volatility, reflecting a small market (in MW terms) spread over a large area. This principle leads to access being firm rather than fixed, and represents a major difference from the SMD.

2.3.5 Transmission Planning and Reliability

Introduction

All electricity markets must be designed to maintain reliability of supply to customers. At the wholesale/transmission level, reliability can be divided into three components:

- *Generation reliability:* sufficient generation capacity to, in principle, reliably supply peak demand: this is achieved by maintaining a capacity margin between aggregate generation capacity and peak demand to cover occasional generator outages.
- *Generation-side transmission reliability:* sufficient transmission capacity to allow the necessary generation to be dispatched and to deliver its output to the market.
- *Demand-side transmission reliability:* sufficient transmission capacity to allow delivery from the market to customer locations.

Obviously the "market" is notional and not located at a particular physical location. So there is no simple, practical delineation between the generation-side and demand-side transmission networks. However, it is a useful concept for thinking about reliability.

Reliability must be maintained through a combination of market and regulatory mechanisms. For the market to drive transmission reliability there must be a link between the decentralised market-based decisions of generators and the centralised regulation-driven decisions of TNSPs. In both the SMD and the OFA, firm access provides the link.

SMD Capacity Markets

As noted earlier, the SMD includes a capacity market. In a capacity market, generators are paid just for being available; they do not need to be dispatched. Capacity is centrally procured, with the objective of maintaining a regulated level of capacity margin. The capacity market financially supports all generators, but is most significant for peaking generators, whose energy market earnings are highly uncertain.

The SMD allows only *firm* generators to participate in the capacity market. Generators become firm by paying a charge that funds the generation-side transmission upgrades needed to ensure that the generator can reliably be dispatched when needed to supply peak demand. Firm generators are also provided with FTRs. In this sense, they are similar to firm generators in the OFA, but capacity market qualification gives them a strong *additional* incentive to go firm.

NEM Reliability Standards

The NEM design does not regulate the capacity margin and so does not need a capacity market to achieve this.[29] Instead, generation reliability is maintained through generators independently deciding how much capacity to build, incentivised by the high peak energy prices on offer when the capacity margin is low.

In the absence of a capacity market, there is no mechanism (currently) for generators to maintain and pay for generation-side transmission reliability. Instead, jurisdictional *reliability standards* place the obligation on TNSPs to provide this transmission capacity, which is funded through TUOS charges.

Thus, the current NEM design contrasts interestingly with the SMD design. The NEM is *more* regulated in relation to generation-side transmission reliability, but *less* regulated in relation to generation reliability.

Reliability under the OFA Design

The OFA design brings the NEM closer to the SMD in relation to reliability. Under OFA, firm generators prompt the generation-side transmission expansion that the TNSP must undertake to maintain FAS. This in turn helps in maintaining reliability.

The incentive to become firm comes from the extreme peak energy prices – which a firm generator has more certainty of receiving – rather than from a capacity market. NEM experience has shown that these peak prices deliver generation reliability. It is unclear whether, under OFA, they would similarly deliver generation-side *transmission* reliability.

Given this uncertainty, it has been decided to maintain existing reliability standards. It is prudent to maintain regulated reliability standards as a safety net should the market fail to deliver reliability. In any case, reform of transmission reliability standards was already under consideration when OFA development commenced, so it made sense to avoid duplicating this work within the OFA project.[30]

Conclusion

There are some similarities, but also some key differences, between OFA and SMD in how they maintain transmission reliability. Critically, the SMD has a capacity market,

---

29 Although a safety net was introduced at NEM commencement, in which the Reserve Trader (AEMO) was able to intervene and purchase capacity reserve where the capacity margin provided by the market is considered insufficient.

30 See: http://www.aemc.gov.au/Markets-Reviews-Advice/Review-of-the-national-framework-for-transmission.

in contrast to the NEM's energy-only market design, so the incentives on generators to drive and fund generation-side transmission development are quite different.

2.3.6 Summary

The key similarities and differences between SMD and OFA are summarised in Table 2.3 below:

Table 2.3 Similarities and differences between SMD and OFA

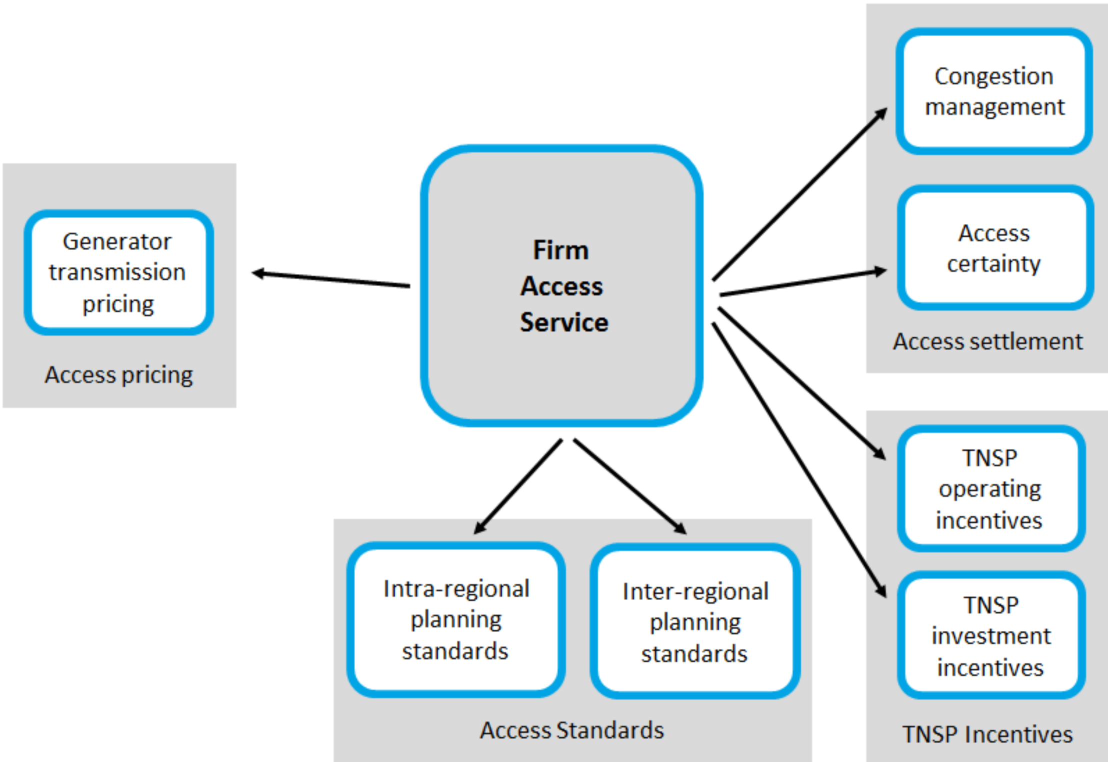
Design element	SMD	OFA
Generation pricing	LMP	LMP
Customer Pricing	LMP or zonal price	LMP at RRN
Settlement balancing	Over the long-run	Within each half-hour
FTR/access firmness	Fixed	Firm
Capacity incentives	Capacity market	Extreme energy prices
Firmness incentive	Participate in capacity market	Access to extreme energy prices

3 OFA Model Overview

3.1 Model Scope

The OFA model is designed to address issues arising on the generation-side of the shared transmission network. These issues all arise because of the way that network access is provided to generators in the current National Electricity Market (NEM) design, and are addressed in the OFA model by the introduction of the firm access service, and the delinking of network access from dispatch. This scope is illustrated in Figure 3.1, below, which shows how the resolution of all of these issues has a common factor: the introduction of firm access service.

Figure 3.1 Transmission issues addressed by the TFR model

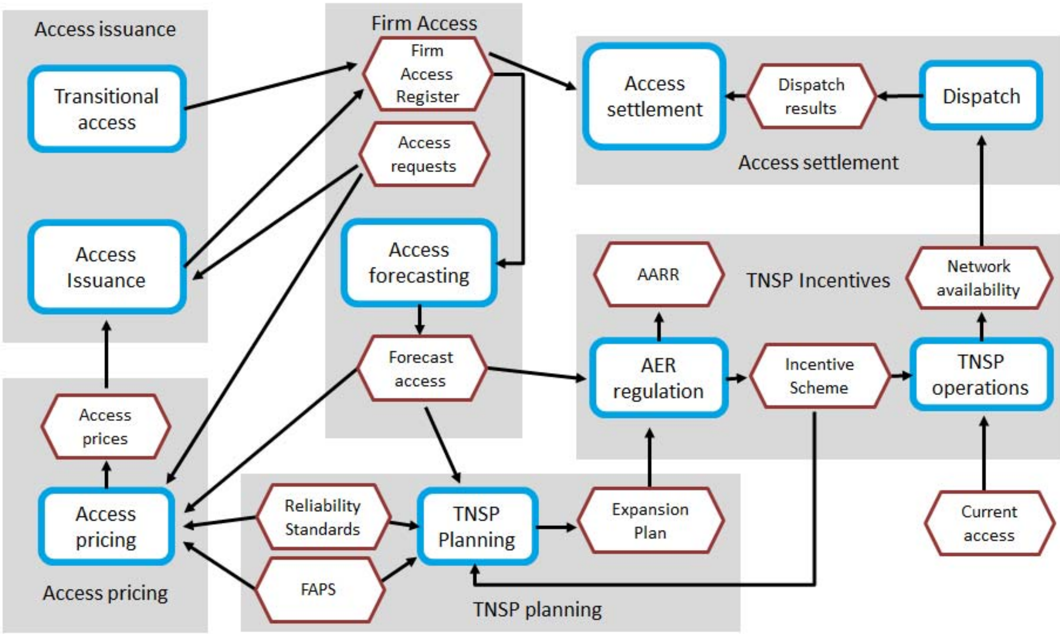


The OFA model does not address, and is not intended to address, transmission issues outside of this scope.

3.2 Model Architecture

The architecture of the OFA model is presented in Figure 3.2, below.

Figure 3.2 OFA model architecture



Although many processes associated with transmission provision and use will change under the OFA regime, this document focuses on five key processes – and one new standard - that are either new or substantially augmented from the current NEM arrangements, as presented in Table 3.1, below.

Table 3.1 Key OFA Model Processes

Process/Standard	Status	Document Chapter
Access Settlement	New	4
Firm Access Standard	New	5
Access Pricing	New	6
Access Issuance	New	7
TNSP Regulation	Augmented	8

These processes are considered in turn in the chapters below.

4 Access settlement

4.1 Overview

Access settlement is the process which *effects* the de-linking of access from dispatch, described in chapter 2. Access is allocated to generators based on their registered access level and their capacity, taking into account the competing access demands of other generators and the fundamental constraint that, to ensure settlement balancing, aggregate access cannot exceed flowgate capacities.

Existing Australian Energy Market Operator (AEMO) settlement calculations and processes are unchanged. Existing settlement payments provide a level of network access *equal* to dispatch level. Therefore, where the access to be allocated to a generator is higher than its dispatch level, the generator *receives* payments *from* access settlements in order to increase its access. On the other hand, where allocated access is *lower* than its dispatch level, a generator makes payments *into* access settlement in order to *reduce* its access.

Access settlement occurs around congested *flowgates*: bottlenecks in the transmission network which are represented by binding constraints on the dispatch process. Typically, there are no more than a handful of congested flowgates in a region in any particular settlement period. So flowgate-based access settlement is not as complex as it first appears.

Similar to generators, directed interconnectors (DICs) currently receive inter-regional access based on their dispatch levels and participate in access settlement so that they can be allocated an access amount that is independent of dispatch.

Access is provided to generators and DICs through the allocation of entitlements on each congested flowgate. Firm generators and firm DICs are given priority in the entitlement allocation process and so enjoy a firmer level of access. Settlement payments to DICs – from regional settlement and from access settlement - are allocated to those market participants that have purchased firm interconnector rights (FIRs).

Generators that provide flowgate support – helping to relieve congestion – are not paid from access settlement and so continue to be paid the Regional Reference Price (RRP) for their output. Therefore, their access continues to be based on dispatch.

There are several practicalities to address for access settlements, including treatment of losses, definition of generators and generator output, and converting 5-minute dispatch values into 30-minute settlement values. There is also a need to define how access settlement operates under abnormal conditions, such as when there are extreme or administered price outcomes.

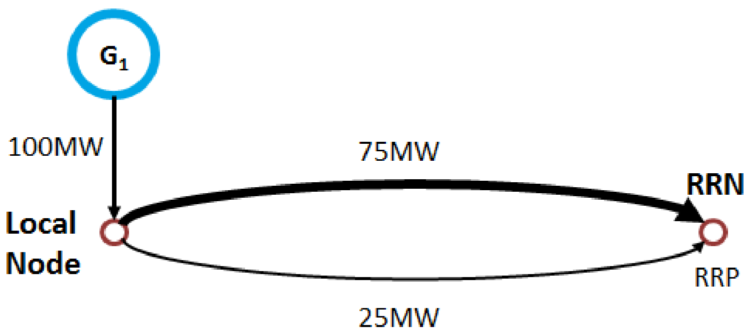
4.2 Design Blueprint

4.2.1 Flowgate Settlement

Flowgates

In the OFA model, the locations in the shared network where congestion may occur are referred to as *flowgates*. In the simple networks discussed in Chapter 2 (presented in Figures 2.3 and 2.4) there is a single flowgate, lying between the two nodes. In a real, meshed network, there are hundreds of flowgates: congestion can potentially occur on any transmission line, as well as across regional or zonal boundaries. Locations where congestion actually occurs are called *congested flowgates*. In the simple network examples, the single flowgate is congested. In a real, meshed network, *several* flowgates may be congested concurrently.[31]

Figure 4.1 Two-path network



Every transmission line and network transformer has a *thermal limit:* the maximum power that can flow through the line before it overheats. Therefore, each of these network elements is a flowgate, referred to as a *thermal flowgate*.

There are limits on power flows that do not relate to the overheating of a particular transmission asset but instead are needed to prevent the power system becoming unstable. These are referred to as *stability flowgates*.

Flowgate Usage

In a meshed network, power flow from a generator to the Regional Reference Node (RRN) will be distributed across multiple paths. Figure 4.1 presents a simplified example where there are just two paths and the power flow is distributed between them in a ratio of 3 to 1. The proportion of the power from a generator that flows through a particular flowgate is referred to as the *participation factor*. In the example,

31 Which may mean a handful of flowgates, but not hundreds. It is only the weakest links in the transmission network which constrain dispatch, meaning that the myriad stronger links cannot become congested.

the participation factors for generator G1 are 75% and 25% on the two paths. Note that the output from the generator is assumed to flow to the RRN. Implicitly there is 100MW of demand at the RRN, although this is not generally shown in the examples.

The amount of power from a particular generator that flows through a particular flowgate, *on its way to the RRN* is referred to as that generator's *flowgate usage*. It is seen that it is simply the product of the participation factor and the generator output:

$$U = \alpha \times G$$

Where:

U = flowgate usage

α = participation factor

G = generator output

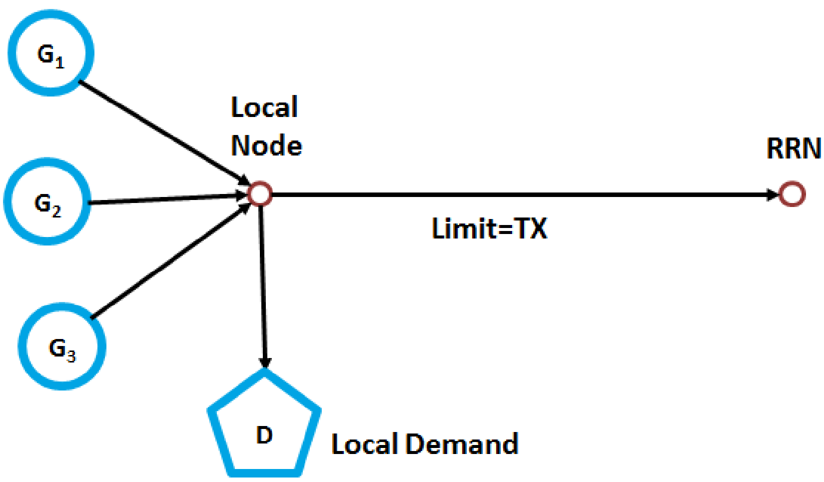
As noted above, every transmission line is a thermal flowgate. Thus, a generator makes use of multiple flowgates. If one or more of these flowgates is congested then its generation may become constrained in dispatch in order to manage the flowgate congestion.

Local Demand

In the simple network examples presented so far, all of the generation output flows through the transmission network to the RRN. When there is *local demand* connected to the local node, some of the output will serve this demand, with the remainder flowing to the RRN.

Figure 4.2 adds local demand to the simple two-node network.

Figure 4.2 Local demand example



The flow through the network is the residual generation after the local demand has been served, ie:

$$\text{Flow} = G_1+G_2+G_3 - D \leq TX$$

Rearranging this inequality gives:

$$G_1 + G_2 + G_3 \quad \leq \quad TX + D$$

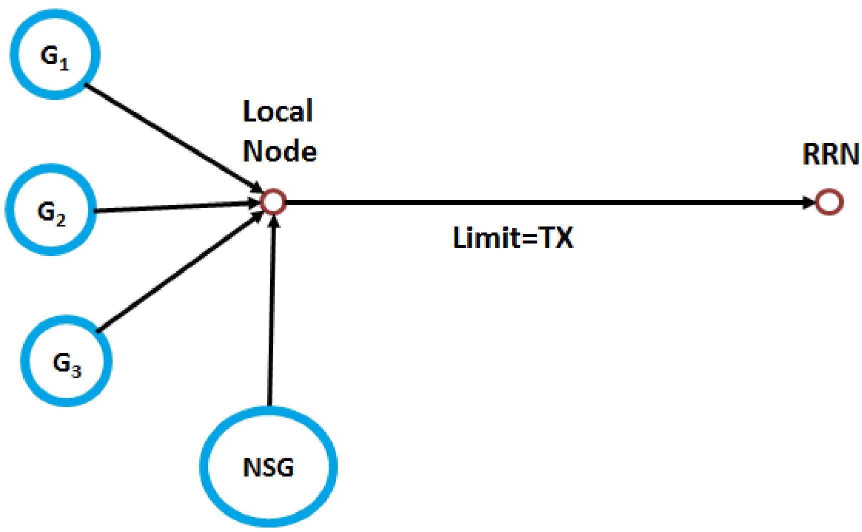
This is equivalent to the situation of no local demand but a larger transmission line with thermal limit TX+D, instead of TX.

In OFA, flowgates are defined in this alternative way: local demand is treated as enlarging the flowgate size and all local generation is considered to use the flowgate, rather than some serving the local demand. The enlarged limit (TX+D in the example) is referred to as the *flowgate capacity*. The physical flow limit, TX, is referred to as the *network capacity*. The difference between these two quantities reflects the level and location of local demand.

Local Non-scheduled Generation

Local non-scheduled generation (NSG) has an equal and opposite effect to local demand. Figure 4.3 replaces local demand with local NSG.

Figure 4.3 Local non-scheduled generation example



In this case the flow inequality becomes:

$$\text{Flow} = G_1+G_2+G_3 + NSG \leq TX$$

Because, like demand, the level of NSG is not controlled by NEMDE, the variable is moved to the RHS of the inequality, which becomes:

$$G_1+G_2+G_3 \leq TX- NSG$$

The flowgate capacity is now the difference between network capacity and the NSG.

Flowgate Entitlements

In the simple examples in chapter 2, for a generator to have access to the RRN it must have access on the flowgate that lies in its path, for which it is credited in access settlement. Similarly, in the general case, where there are multiple paths and flowgates, for a generator to have access it must be credited in access settlement on all of the congested flowgates that lie on paths to the RRN. The amount of this credit on each flowgate is referred to as the flowgate *entitlement*. As in the simple case, the entitlement is equal to what the flowgate usage would be if the generator was dispatched at its access level. Using the formula for usage, the entitlement is therefore:

$$E = \alpha \times A$$

That is to say, if a generator is allocated an entitlement on every flowgate according to the above formula then it obtains an access level A to the RRN.[32]

Flowgate Prices

As described in chapter 2, network access means a payment based on the formula:

$$\text{Payment\$} = A \times (RRP-LMP) \qquad (2.1)$$

In the simple two-node example, the price difference, RRP-LMP, represents the value, at the margin, of the flowgate capacity. If the flowgate capacity could be increased by 1MW then the marginal, constrained-off generator at the local node could have its output increased by 1MW, at a cost of LMP.[33] Its output would displace the marginal generator at the RRN, saving RRP. The net saving is the difference between RRP and LMP. The marginal value of flowgate capacity is referred to as the *flowgate price (FGP)*.

In the general situation of a meshed network with multiple flowgates, every flowgate has an associated price, defined similarly as the marginal value of the flowgate capacity. However, the formula for calculating flowgate prices are complex, as discussed in more detail in appendix A.1. Fortunately, flowgate prices are already calculated during the dispatch process.

If a flowgate is uncongested, adding to its capacity simply increases the amount of unused capacity: it will not change the dispatch outcome and there is therefore no associated cost saving. Therefore the flowgate price is zero.

On the other hand, if a flowgate is congested, it will be causing some generation to be constrained and replaced by more expensive generation. Thus, if the flowgate capacity

32 This is demonstrated mathematically in appendix A.1.

33 The marginal generator's offer price must be LMP, since it sets the price at the local node.

could be increased, there would be some cost saving. Therefore, the flowgate price is *greater* than zero. Therefore, flowgate prices are *never* negative.

Settlement Algebra

The discussion above explains how the simple network examples in chapter 2 can be generalised to a complex, meshed, real-world network. These generalisations are summarised in Table 4.1 below.

Table 4.1 Settlement variables and their equivalents in the two-node model

Variable	Acronym	Description	Value in simple intra- or inter-regional models
flowgate price	FGP	the marginal value of flowgate capacity	RRP-LMP or RRP_M-RRP_X
flowgate usage	U	the amount of a generator's output, or a DIC flow, that flows through the flowgate	G or I
flowgate entitlement	E	the amount of network access that a generator or DIC is allocated through the flowgate	A or A_{IR}
flowgate capacity	FGX	the maximum aggregate flowgate usage which the flowgate can accommodate	TX

Access settlement takes place separately at every congested flowgate. The generalised variables are used. Recall that, in the simple example presented in chapter 2, access settlement is given by the formula:

$$Pay\$ = (A-G) \times (RRP-LMP)$$

Using the generalisations in the table above, the general access settlement equation becomes:

$$Pay\$ = (E - U) \times FGP \qquad (4.1)$$

Settlement Balancing

Recall from section 2.3.4 that it is required that access settlement balances. In the simple example, a sufficient condition for this is that:

$$A_T = TX$$

Where:

A_T = total allocated access (including any inter-regional access)

TX = transmission capacity

Using the generalisations, this condition becomes:

$$E_T = FGX \qquad (4.2)$$

Where:

E_T is the total of all entitlements (including any inter-regional entitlements) on a flowgate.

Equation 4.2 ensures that access settlement balances in relation to a particular flowgate. So long as the equation holds for all flowgates then access settlement will balance in aggregate.

In the OFA model, equations (4.1) and (4.2) apply to *all* congested flowgates in all meshed networks. They are the basic building blocks of access settlement.

Constraint Equations

As part of the current dispatch process, AEMO prepares a set of constraint equations[34] that reflect potential constraints on dispatch caused by limitations in the transmission network.[35] The distinction between transmission and non-transmission constraints is discussed further in section 4.3.2. These constraint equations are fed into the NEM Dispatch Engine (NEMDE), which then finds an economic (ie, lowest cost) dispatch solution which complies with these constraints.

The constraints are all *linear*. This means that all *intra*-regional constraints, those that do not involve interconnectors[36] take the form:

$$\alpha_1 \times G_1 + \alpha_2 \times G_2 + \ldots + \alpha_N \times G_N \leq RHS$$

Where:

α_1, α_2 etc are fixed coefficients

G_1, G_2 etc are generation dispatch targets

RHS is the "right hand side" amount which is independent of generation dispatch

Since each constraint equation relates to a potential limit on the transmission system, it represents a flowgate: there is a one-to-one correspondence between transmission constraint equations and flowgates.

34 Strictly speaking, these are inequalities rather than equations.

35 AEMO also prepares constraints unrelated to transmission, but these are not relevant to access settlement.

36 Inter-regional constraints, those involving interconnectors, are discussed in section 4.2.3.

Comparing this equation to the discussion of earlier discussion of flowgates, it will be seen that:

- the coefficients in the constraints equations are the participation factors;[37]
- the RHS is the flowgate capacity[38]; and
- the individual terms on the left-hand side, $\alpha_i \times G_i$, represent the flowgate usage of each generator.

Substituting the dispatch constraint variables with the corresponding flowgate variables, the constraint equation can be re-written as:

$$U_T \leq FGX$$

Where:

U_T is total usage

When NEMDE runs, as well as calculating dispatch targets, it also calculates a "marginal value" for every binding constraint which is the same as the FGP discussed previously. Thus, all of the variables needed for access settlement, except for entitlements, are already prepared or calculated as part of the existing dispatch process.

Settlement Architecture

Access settlement calculates the amounts payable to or from each generator (and DIC) by applying the two fundamental equations derived above:

$$Pay\$ = (E-U) \times FGP$$

$$E_T = FGX$$

It does this using three basic processes, presented in Table 4.2 below.

37 Actually, this statement is not strictly true because, when AEMO formulates the constraint equations, it scales up the coefficients by a common factor such that the largest coefficient equals unity, and scales the RHS by the same factor, so that the constraint equation is left unaffected. For example, a constraint $0.25 \times G \leq FGX$ would be scaled up by AEMO to $1 \times G \leq 4 \times FGX$. The *true* participation factor is 0.25 but the constraint coefficient is 1. This scaling does not affect access settlement payments. Although settlement *volumes* are scaled up as a result, settlement prices (ie, FGP) are also scaled *down* by the same factor. Therefore, the constraint coefficients can be used in access settlements as *though* they are true participation factors. In the context of access settlement, "participation factor" will always refer to the constraint coefficient, not to the (true) participation factor.

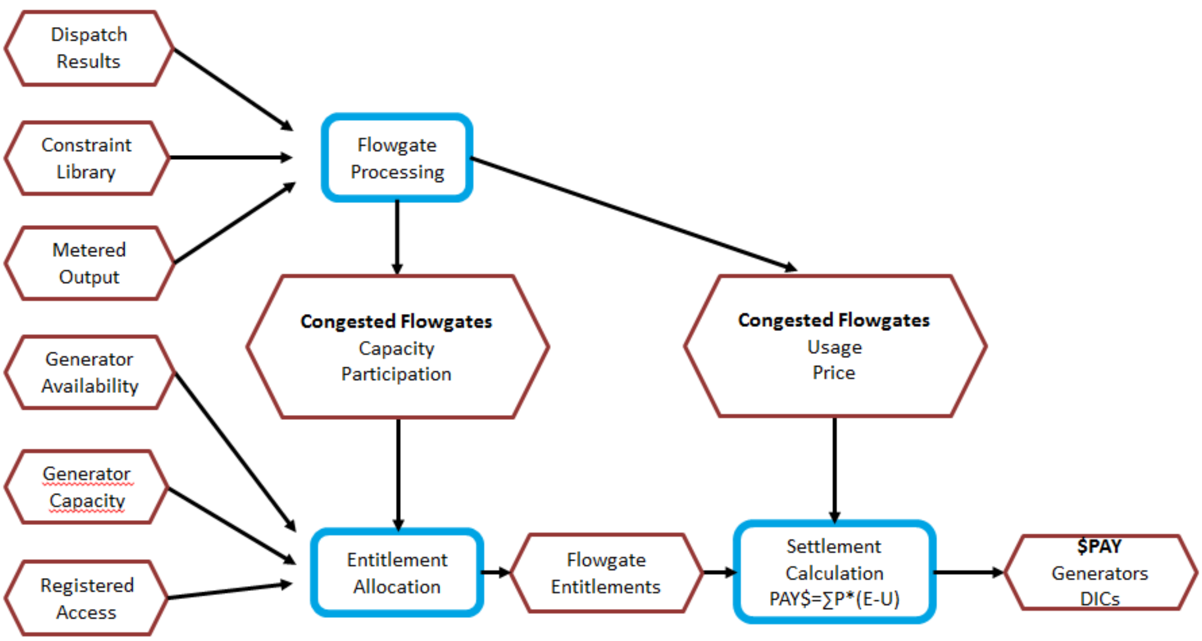
38 Again, the RHS in fact represents the scaled up flowgate capacity. In the context of access settlement, "flowgate" will always refer to the constraint RHS, not to the (true) flowgate capacity.

Table 4.2 Access settlement processes

Process	Description
Flowgate Processing	Determines price, capacity, usage and other relevant variables for each congested flowgate
Entitlement Allocation	Allocates the flowgate capacity between generators, ensuring that total entitlement equals flowgate capacity
Settlement Calculation	Applies the formula Pay$ = FGP x (E-U) to each generator at each congested flowgate

The linkages between these processes and existing NEM databases are shown in Figure 4.4 below.

Figure 4.4 Access settlement processes



Flowgate processing extracts, from existing dispatch databases, the variables relevant to access settlement, which are discussed in the previous section. Entitlement allocation determines entitlements for each generator and DIC on each flowgate and is described in section 4.2.4. The Settlement Calculation process then uses the prices and volumes determined in these other processes to calculate payment amounts for each generators and DIC.

As noted previously, every transmission line[39] is a thermal flowgate and there are other, stability flowgates. Thus there may be thousands of flowgates across the NEM. AEMO prepares constraint equations representing all of the flowgates that could potentially become congested: these typically number in the hundreds for any particular dispatch run. However, it is rare for more than ten or twenty flowgates to be congested concurrently.

Therefore, whilst conceptually complex, access settlement is practically relatively straightforward.[40] Much of the data needed is already available in dispatch databases. In a settlement period, settlement calculations occur only on a handful of the myriad constraints that are contained in AEMO constraint libraries.

Local Prices

Flowgate settlement appears to be quite different to the simple, LMP-based access settlement described in section 2.2.3. But, in fact, the mathematical relationship between flowgate prices, participation factors and LMPs means that the aggregate payment to a generator under flowgate settlement can similarly be expressed in terms of LMP and RRP.

$$\text{Access Pay\$} = (\text{RRP-LMP}) \times (\text{A-G})$$

In this case, the *effective access level*, A, is a weighted-average of the entitlements provided to the generator on the various congested flowgates. This result is demonstrated mathematically in appendix A.1.4.

Therefore, in conjunction with regional settlements, a generator is paid:

$$\text{Total Pay\$} = \text{LMP} \times \text{G} + (\text{RRP - LMP}) \times \text{A}$$

As discussed in section 2.2.1, this equation ensures that the no-regret dispatch principle applies whenever:

- a generator bids at or above cost; and
- access (ie, entitlements) are independent of dispatch.

4.2.2 Flowgate Support

Flowgate Access and Flowgate Support

Participation factors can be positive or negative. A generator with a *positive* participation factor has positive usage on a flowgate. Thus, flowgate congestion may be managed in dispatch by reducing the output – and therefore the flowgate usage - of such positive-participation generators. Such generators are said to be *constrained-off*. It is for these generators that the OFA model is designed. By separating access from dispatch, the model ensures that generators that are constrained-off do not lose access

---

39 And also every network transformer.

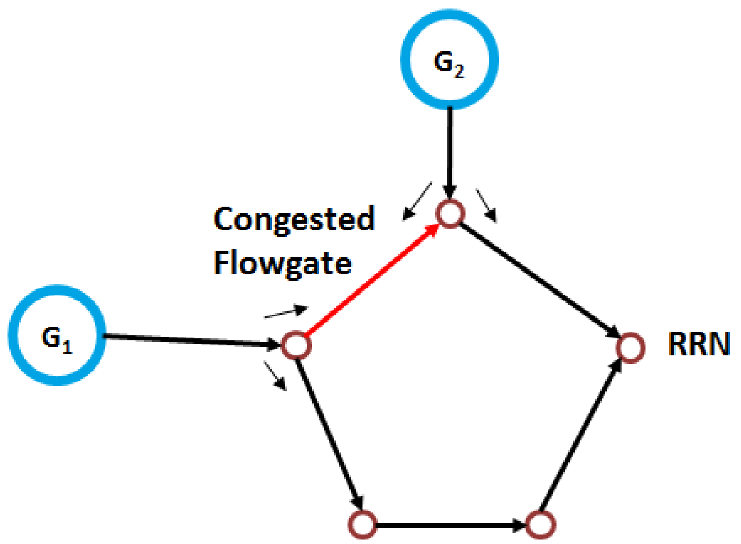
40 Although, as with any settlement process, there are a number of practicalities to work through: these are discussed in section 4.2.5.

as a result. These positive-participation generators are therefore referred to as *flowgate access generators.*

On the other hand, a generator with a *negative* participation factor has negative usage and so flowgate congestion may be managed in dispatch by *increasing* the output, and thus reducing the flowgate usage[41] of negative-participation generators. Because they help to relieve congestion, they are referred to as *flowgate support generators.*[42]

For thermal constraints, negative participation, and flowgate support, occur only on meshed networks and so did not arise in the simple, radial network examples presented in Chapter 2. Figure 4.5 illustrates how flowgate support and flowgate access generators are generally located on a loop that contains both the congested flowgate and the RRN. Given the location of the congested flowgate, G_1 is a flowgate access generator whose output exacerbates the congestion. G_2 is a flowgate support generator whose output helps to relieve the congestion.

Figure 4.5 Loop flow constraint



If there is no loop, then there will be no flowgate support.[43]

Because they help to relieve congestion, flowgate support generators have a high value in dispatch. This is reflected in a LMP that is higher than the RRP. As discussed in section 2.2.1 generators are dispatched when their local price is higher than their offer

41 Ie, making it more negative.

42 It is possible that a generator participates in two congested flowgates: one with positive participation and one with negative participation. A generator subject to such mixed constraints would be regarded as a flowgate access generator on one flowgate and as a flowgate support generator on the other. Mixed constraints are discussed further in appendix A.1.

43 At least for thermal flowgates. This may not be true for stability flowgates.

price. Thus, a flowgate support generator may be constrained-on: dispatched despite the fact that the RRP is below its offer price.

Settlement of Flowgate Support Generators

Access settlement payments to flowgate support generators are set at zero. These generators therefore continue simply to be paid RRP, under existing regional settlement. The reason for this approach is discussed in section 4.3.3.

Recall the basic formula for access settlement:

$$\text{AccessPay\$} = (E - U) \times FGP$$

For access settlement payments on congested flowgates to be zero, we must have:

$$E = U$$

Therefore, entitlements for flowgate support generators are set equal to their usage on every congested flowgate. Flowgate support generators have negative usage and so they are allocated negative entitlements.

Flowgate Support Amount

To ensure access settlement balances, it is required that:

$$E_T = FGX \tag{4.3}$$

Where E is *total* generator entitlements. This can be separated into two components:

$$E_T = E_A + E_S \tag{4.4}$$

Where:

E_A = total entitlements to flowgate *access* generators; and

E_S = total entitlements to flowgate *support* generators

Since flowgate support generators have negative entitlements individually, the aggregate, E_S is also negative. *Flowgate support* (FGS) is a positive quantity which is defined as the *negative* of E_S:

$$FGS = -1 \times E_S \tag{4.5}$$

The three equations above can now be combined and rewritten as:

$$E_A = FGX + FGS$$

The amount of total entitlements that can be allocated to access generators is referred to as the *effective* flowgate capacity, EFGX. Hence:

$$EFGX = FGX + FGS \qquad (4.6)$$

Thus, flowgate support generators facilitate an increase in the level of entitlements allocated to flowgate access generators; they support increased access.

For example, suppose a constraint takes the form:

$$G_1 + G_2 - G_3 < 100MW$$

G3 has a participation factor of minus one and so is a flowgate support generator. If G_3=50MW (say) then G3 provides flowgate support of 50MW. This makes the *effective* flowgate capacity equal to 150MW, which will be allocated between G1 and G2.

4.2.3 Interconnectors

Inter-regional Flowgates

Constraints which involve interconnectors (they may also contain generators) will be referred to as *inter*-regional constraints. Each intra- or inter-regional constraint gives rise to an intra-regional or inter-regional flowgate, respectively.

The general form of an inter-regional constraint is:

$$\alpha_1 \times G_1 + \alpha_2 \times G_2 + \ldots + \alpha_N \times G_N + \alpha^{IC}{}_1 \times IC_1 + \alpha^{IC}{}_2 \times IC_2 + \ldots + \alpha^{IC}{}_M \times IC_M \leq RHS$$

Where:

α^{IC} is the participation factor for an interconnector

IC is the flow on an interconnector

In NEMDE, α^{IC} may be positive or negative. The sign indicates whether the constraint limits the amount of inter-regional flow north or south, respectively.[44] Unlike with generators, it does not establish whether or not the interconnector is providing flowgate support.

DIC Participation

Inter-regional constraints refer to interconnectors but, as discussed in section 2.2.2 it is *directed* interconnectors (DICs) that participate in access settlement. Therefore, whenever an interconnector participates in a flowgate, it must be decided which of the two DICs associated with that interconnector is the flowgate participant (ie, the DIC that participates in access settlement).

In normal circumstances, the DIC whose increased flow *exacerbates* the congestion is deemed to be the flowgate participant. In exceptional circumstances, discussed in

44 This is based on the sign convention that AEMO uses for interconnector flow: a positive amount indicates a flow in a northerly or westerly direction.

section 4.2.4, the other DIC participates: the one whose increased flow *relieves* congestion.

For example, suppose that $\alpha^{IC} > 0$ in a particular flowgate. This means that increased northerly interconnector flows will exacerbate congestion. Therefore, in this case, the *northerly* DIC is the participant.

On the other hand, if $\alpha^{IC} < 0$ in a flowgate, the southerly DIC is the participant.

This is summarised in the table below.

Table 4.3 DIC Participation

Sign of α^{IC}	Participating DIC	DIC participation factor	DIC flow	DIC usage
positive	northerly	equals α^{IC}	equals IC	IC × α^{IC}
negative	southerly	equals -1 x α^{IC}	equals -1 x IC	IC x α^{IC}

Note that:

- the participation of the DIC is always positive; and
- the flowgate usage of the DIC is always the same as the interconnector flowgate usage: since DIC flow can be positive or negative[45], usage can also be positive or negative.

DIC Settlement

In section 2.2.3, the access settlement payment for a DIC was defined for a simple, three-node network and a single inter-regional constraint:

$$DIP\$ = I \times (RRP_M - RRP_X) + (A^{IR} - I) \times (RRP_M - RRP_X)$$

Where:

DIP\$ is the total settlement payment to the DIC

RRP_M and RRP_X are the RRPs in the importing and exporting regions, respectively

A^{IR} is the amount of inter-regional access provided to the DIC

I is the DIC flow

45 A negative flow is referred to as a counter-price flow. This situation is discussed further in section 4.3.4.

As was the case for generator settlement, the formula for flowgate settlement is derived from this basic form by substituting:

- $(RRP_M - RRP_X)$ with FGP
- A^{IR} with E^I: the flowgate entitlement allocated to the DIC
- I with U^I: the DIC flowgate usage

The equation then becomes:

$$DIP\$ = U^I \times FGP + (E^I - U^I) \times FGP \qquad (4.7)$$

The first term in equation 4.7 is paid out of the IRSR from regional settlements, as discussed in the next section. The second term is calculated and paid in access settlements as discussed in section 4.2.1.[46] Settlement balancing requires that DIC entitlements are allocated out of the flowgate capacity, ie:

generator entitlements + DIC entitlements = FGX

Allocation of IRSR between DICs

Where an interconnector participates in several congested flowgates concurrently, the corresponding IRSR must fund multiple payments. The mathematical relationships between the RRPs, FGPs and participation factors mean that the IRSR is always sufficient[47] to fund these payouts[48], ie:

$$IRSR\$ = U^I{}_1 \times FGP_1 + U^I{}_2 \times FGP_2 + \ldots U^I{}_N \times FGPN \qquad (4.8)$$

Depending upon the direction of these various flowgates, the IRSR could be allocated in its entirety to one or other of the two DICs on the interconnector, or it could be shared between them.

4.2.4 Entitlement Allocation

Overview

The following principles have been established in relation to flowgate entitlements:

1. access settlement payments for each generator or DIC on each congested flowgate are based on the difference between flowgate entitlements and flowgate usage;

---

46 Note that the equation (4.1) presented in that section applies equally to generators and DICs.

47 Ignoring losses as usual.

48 This is demonstrated in appendix A.2.

2. flowgate entitlements must be allocated between flowgate access generators and DICs such that total entitlements on a flowgate equals effective flowgate capacity; and

3. firm generators and DICs (those who receive a firm access service from their TNSP) are given priority allocation of entitlements.

This section describes how these principles are to be applied within access settlement in order to allocate flowgates entitlements to flowgate access generators and DICs. Recall that the allocation of (negative) entitlements to flowgate support generators has already been described, in section 4.2.2.

Target Access

For a generator or DIC, *target access* is the maximum amount of access that it will be allocated. Target access is made up of firm and non-firm components. Formulae for these targets are defined in Table 4.4.

Table 4.4 Target Access Amounts

Target Access Component	Formula
Generator Target Firm Access	Lower of registered access and capacity[49]
Generator Target Non-firm Access	Amount (if any) by which offered availability exceeds firm access amount
DIC Target Firm Access	Registered Access
DIC Target Non-firm Access	Zero

For example, a 300MW generator that purchases 500MW of firm access will only be credited with 300MW's worth of access in access settlement. The reasons for capping the target access in this way are discussed in section 4.3.5.

Interconnector flows are not inherently constrained by a capacity limit in the same way as generators; they are only ever constrained by transmission constraints. Therefore, a capacity limit is *not* applied to DIC target access.

Some numerical examples of generator target firm access and target non-firm access are shown in Table 4.5 below.

49 Capacity is determined empirically as the maximum output of that generator in a settlement period, over the previous two years.

Table 4.5 Numerical examples of generator target firm and non-firm access

Generator Type	Registered Access	Capacity	Availability	Target Firm Access	Target Non-Firm Access
Super-firm	500	300	300	300	0
Firm	300	300	300	300	0
Part-firm and available	200	300	300	200	100
Part-firm and unavailable	200	300	0	200	0
Non-firm and available	0	300	300	0	300
Non-firm and unavailable	0	300	0	0	0

Note that the generator availability only affects the target access of part-firm and non-firm generators.

Target Entitlements

Recall from section 4.2.1 that to provide a generator with an access level *A* it must be provided, on each congested flowgate, with an entitlement given by the formula:

$$E = \alpha \times A \tag{4.9}$$

These amounts are referred to as target entitlements. Where generators and DICs have firm and non-firm target access amounts, they have corresponding target entitlements:

Target firm entitlement = α × target firm access

Target non-firm entitlement = α × target non-firm access

Equation (4.9) is another fundamental building block for access settlements. The target entitlements that it defines are calculated dynamically: as congestion arises at different flowgates, the relevant participation factors are extracted from the corresponding NEMDE constraints and target entitlements are then automatically calculated for each participating generator.

A numerical example demonstrating the calculation of target entitlements is presented in appendix A.3.

The aggregate of the target firm entitlements is referred to as the target flowgate capacity (TFGX). This level of flowgate capacity is sufficient to provide the target firm entitlements to all generators and DICs on a flowgate. If it is provided on every congested flowgate, then generators and DICs will receive their target firm access. TFGX is another key concept in the OFA design, since it forms the basis for the Firm Access Standard.

Entitlement Scaling

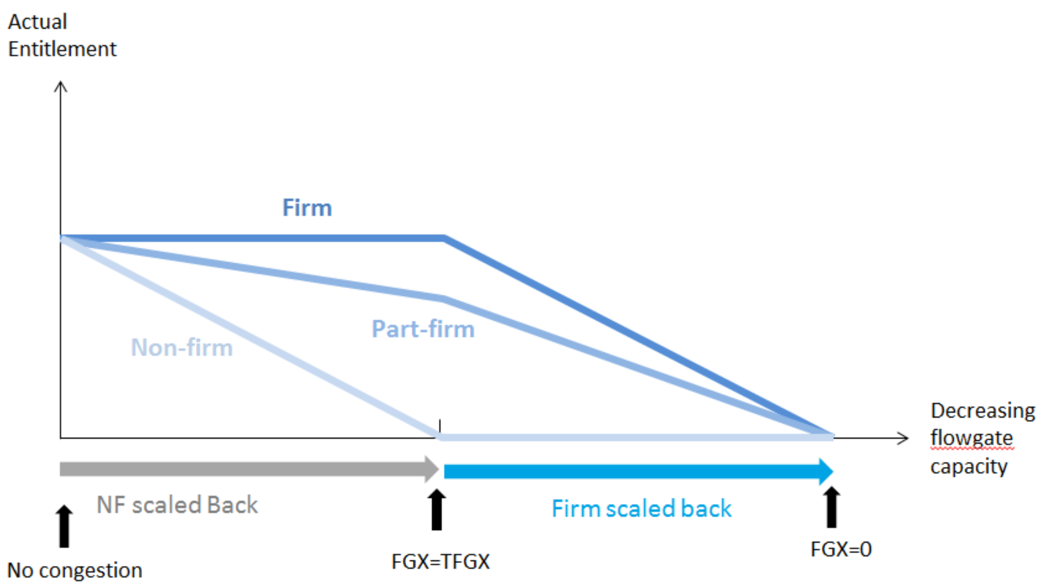
The aggregate of all target entitlements (firm and non-firm) on a congested flowgate will typically *exceed* the effective flowgate capacity, meaning that not *all* entitlement targets can be achieved. An entitlement scaling algorithm is used to determine *actual entitlements* on a flowgate, based on the following principles:

- total actual entitlements must equal effective flowgate capacity;
- actual entitlements are non-negative and do not exceed target entitlements;
- actual entitlements are proportional to target entitlements; and
- firm entitlements are only scaled back when non-firm entitlements have already been scaled back to zero.

The algorithm calculates, for each congested flowgate, a *firm scaling factor* and a *non-firm scaling factor*, applied to firm and non-firm entitlements, respectively. The detailed algebra for determining the scaling factors and entitlements, together with a numerical example, is presented in appendix A.3.

Figure 4.6, below, illustrates the level of entitlements that would be allocated to generators in these three access categories under decreasing levels of flowgate capacity. For simplicity, these generators are each assumed to have identical capacities and participation factors and to be fully available.

Figure 4.6 Entitlement scaling 3 categories



Non-Firm Inter-regional Entitlements

On inter-regional flowgates, it is possible that there will be some residual flowgate capacity even after all firm and non-firm targets entitlements have been fully met. This residual is allocated to any participating directed interconnectors as non-firm actual entitlements. Where there are two or more DICs participating in the flowgate, the entitlements are allocated pro rata to DIC "capacity".[50]

For example, consider a radial inter-regional flowgate (ie, participation factors equal 1) shared between a 1,000MW firm generator and a non-firm directed interconnector. The generator simply has a firm target entitlement of 1,000MW; the interconnector has zero target entitlements. If the flowgate capacity is 1,200MW, say, all target entitlements can be fully met (with 1000MW allocated to the generator and 0MW allocated to the DIC) and 200MW of flowgate capacity remains. This 200MW is allocated to the interconnector as a non-firm actual entitlement.

Flowgate Support Interconnectors

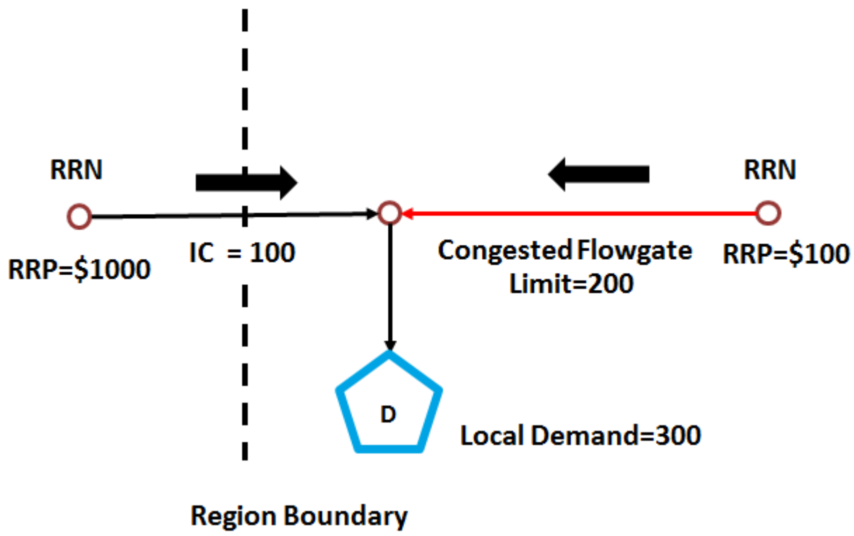
Two principles that have been applied to entitlement allocation are:

- entitlements for DICs and flowgate access generators are non-negative; and
- the aggregate of these entitlements must equal the effective flowgate capacity.

50 Similar to generators, DIC capacity is defined as the maximum flow over a specified recent historical period. Unlike with generators, DIC do not have any inherent capacity limit, so the terminology is used for convenience rather than accuracy.

It will be seen that these two principles conflict when EFGX is negative. Since flowgate capacity is related to transmission capacity it may seem odd to contemplate a negative value. But, in fact, negative FGX can arise when there is local demand connected on an interconnector, as illustrated in Figure 4.7.

Figure 4.7 Negative flowgate capacity



In this figure, the flow through the congested flowgate is a combination of the local demand and the interconnector flow. This may not exceed the transmission capacity:

$$300 - IC \leq 200$$

Where:

IC = easterly flow on interconnector

Rearranging this equation to a standard form gives:

$$-IC \leq -100$$

The IC must be dispatched to an amount of at least 100MW, easterly, to supply the local demand (or to relieve the congestion, depending upon how you look at it).

Note that, in this case, the local demand is located *downstream* of the flowgate and causes a *reduction* in flowgate capacity. This is the opposite of the example presented in section 4.2.1 in which the local demand was *upstream* of the flowgate and caused an *increase* in flowgate capacity.

So, in this example, FGX is minus 100MW. There is no flowgate support generation, so the EFGX is also minus 100MW. The flowgate is congested and so flowgate usage is also minus 100MW.

In the light of the principle described in section 4.2.3, the westerly DIC (with positive participation in the flowgate) would normally be deemed to participate in this flowgate for the purposes of access settlement. However, whenever there is negative EFGX, an exception is made and the opposite DIC instead participates: the easterly DIC in this example; the negative participating DIC, in general.

A negatively participating DIC is treated as a *flowgate support interconnector*: analogous to a flowgate support generator. It is allocated a (negative) entitlement equal to its usage and provides a flowgate support amount of:

Flowgate support = -1 × usage = +100

This flowgate support adds to the effective flowgate capacity and removes the problem of negative EFGX.

EFGX = FGX + FGS = -100 + 100 = 0

Therefore, in this example:

- -100MW of entitlement is allocated to the easterly DIC, as a flowgate support interconnector;
- no flowgate access entitlements are allocated: since EFGX=0; and
- total entitlements = -100 = flowgate capacity.

In summary, DICs only provide flowgate support when this is necessary to avoid a negative EFGX. With the inclusion of flowgate support interconnectors, EFGX can never be negative.[51]

Allocation of DIC payment to NEM Participants

Section 4.2.3 explained how, for each congested inter-regional flowgate, an amount is paid to each participating DIC according to the formula:

$DIP\$ = E^{I} \times FGP$

The DIC is a notional settlement entity that is not owned by any NEM participant. Thus, the DIC payment needs to be allocated to NEM participants. The principle for this allocation is that payments relating to:

- *firm entitlements* are passed on to the relevant FIR holders;
- *non-firm entitlements* are passed onto the TNSP in the importing region; and
- *flowgate support entitlements* are passed onto the TNSP in the importing region.[52]

51 This is demonstrated in appendix A.3.

52 For reasons discussed in section 4.3.4.

Payments to FIR holders are allocated pro rata to each NEM participant's volume of FIRs held. Detailed algebra relating to FIR payments is presented in appendix A.2.2.

4.2.5 Practicalities

Measuring Generation

Access settlement requires various generator quantities to be measured: output, availability and capacity. These measurements can relate to:

- *Sent-out* quantities: the amount of power that is transferred into the shared transmission network.
- *As-generated* quantities: the amount of power that is generated by a power station.
- *Metered* quantities: the amount of energy each half hour as measured by the existing settlement metering.

For conceptual and practical reasons, OFA in general – and access settlement in particular – uses sent-out quantities, since these relate to generators' use of the shared transmission network. These reasons are presented in section 4.3.10.

However, this creates some practical difficulties to be overcome, since existing data is commonly available only on an as-generated or metered basis.

Differences between these approaches stem from the treatment of auxiliary load: electricity that is consumed by a power station as part of the generating process. The calculation of sent-out quantities depends upon the configuration of a power station: the topology and locations of the meters and the auxiliary load connections. Since power station configurations vary widely, the calculation process is rather complex. It is described in detail in appendix B.1.

Generator Entity

This chapter refers to "generators" in the context of access settlement. This term requires a precise definition. In dispatch, and in the associated constraint equations, generators are specified at the level of entities referred to as dispatchable units.[53] In OFA, a generator entity is formally referred to as an access unit. Access units must be defined in such a way that their individual outputs and participation factors are straightforward to determine.

53 The formal name used by AEMO is Dispatch Unit Identifier or DUID.

Typically a dispatchable unit is simply an individual physical generating unit at a power station, but in some cases it is a group of generating units, possibly across multiple power stations.[54]

Since dispatchable units appear in the left-hand side of constraint equations, their participation factors are readily available. However, in some metering configurations, multiple dispatchable units share a revenue meter. In this situation it is not straightforward to determine sent-out output quantities for each dispatchable unit. Where they share a meter, dispatchable units must also share a connection point to the transmission network and so must have common participation factors.

In the light of this, an access unit is defined to be:

- a single dispatchable unit, where this has dedicated a revenue meter; or
- a group of dispatchable units, where these share a revenue meter.

Generally, when “generator” is referred to in this paper, the meaning of “access unit” is intended. However, in some cases, “generator” may refer to a generating company that owns a portfolio of access units. The context should make the distinction clear.

Some examples of dispatchable unit and access unit definitions under different metering and connection configurations are presented in appendix B.1.

Transmission Losses

For simplicity of exposition, transmission losses are ignored in the above description of access settlement. However, transmission losses are a fundamental aspect of regional settlement. Access settlement needs to align with this by also incorporating losses.

Transmission losses within a region are represented in regional settlement by intra-regional *marginal loss factors* (MLFs): coefficients that are calculated annually for every node on the shared transmission network.[55] These loss factors are incorporated into access settlement in two ways.

First, all *usage* amounts are adjusted for losses so:

$$U = MLF \times \alpha \times G$$

Where:

MLF is the marginal loss factor for the relevant generator

This adjustment ensures that the no-regret dispatch principle continues to hold when losses are factored into dispatch and pricing.

54 Such a group is sometimes referred to as an *aggregated* generating unit.

55 And also for distribution networks, although the distribution loss factors represent average losses rather than marginal losses.

Second, all target access amounts are also adjusted for losses so:

Target Firm Access = MLF × (Lower of agreed access and capacity)

Target Non-firm Access = MLF × (Amount if any by which availability exceeds firm access amount)

Unlike the usage adjustment, this adjustment to access targets is not critical to maintaining the fundamental OFA principles. However, on balance it is considered to provide benefits to generators and TNSPs. This is discussed further in section 4.3.9.

Inter-regional MLFs are applied in regional settlements to pricing and dispatch on interconnectors. However, entitlements and usages for DICs are not loss adjusted.

These inter-regional loss factors affect the amount of the IRSR. Recall from section 4.2.3 that the IRSR is allocated to DICs participating in congestion inter-regional flowgates according to the formula:

Allocation from IRSR = FGP × U

The aggregate of these allocations on an interconnector is commonly referred to as the *congestion rent*. In the final subsection of section 4.2.3 it was asserted that, ignoring losses, the congestion rent equals the IRSR. But, when loss adjustments are made, the IRSR amount changes by an amount referred to as the **losses residue**. Thus, by definition:

IRSR = congestion rent + losses residue

After the allocations to DICs described earlier, the losses residue remains and is allocated to the TNSP in the importing region, based on the interconnector flow direction.[56] Where there are parallel interconnectors, the allocation is based on the flow direction of the interconnectors *combined.* Currently, the *entire* IRSR is allocated using this method, so the introduction of OFA makes no change to the allocation of the losses residue portion of the IRSR.

Calculation of Effective Flowgate Capacity

Effective flowgate capacity (EFGX) must be calculated in access settlement as an input to the entitlement scaling process. There is a straightforward way to calculate EFGX that avoids the need to explicitly calculate the constraint RHS. Recall that, for a congested flowgate:

Total entitlements = total usage

56 It is not possible to allocate it according to the direction of congestion. Losses residue will accrue when there is no congestion and also when there are mixed constraints on an interconnector. In both situations, the direction of congestion is undefined.

These totals can be divided into components associated with DICs, flowgate access generators and flowgate support generators:

$$E^T = E^I + E^A + E^S$$

$$U^T = U^I + U^A + U^S$$

Where:

E^T = total entitlements

E^I = total DIC entitlements

E^A = total flowgate access generator entitlements

E^S = total flowgate support generator entitlements

U^T = total usage

U^I = total DIC usage

U^A = total flowgate access generator usage

U^S = total flowgate support generator usage

Recall that, for flowgate support generators, entitlements are set equal to usages, and so:

$$E^S = U^S$$

And recall also that EFGX is allocated between flowgate access generators and DICs:

$$E^A + E^I = EFGX$$

Putting together these various equations we have:

$$EFGX = E^A + E^I = E^T - E^S = U^T - U^S = U^A + U^I$$

Therefore, EFGX is simply the aggregate usage of all flowgate access generators and DICs. Note that these usages are loss-adjusted, as described in section 4.3.9. Note also that the EFGX calculated in this way excludes any flowgate support from interconnectors.

Thirty-minute Settlement

The settlement period for access settlement is a trading interval (30 minute period), which is the same as for existing regional settlement. However, many of the dispatch variables – such as usage and flowgate capacity – are calculated by NEMDE each dispatch interval (5 minute period). These quantities are converted in access settlement to 30 minute equivalents, using averaging. Alternative approaches are discussed in section 4.3.7.

Extreme Prices

In regional settlements, limits are placed on RRPs: they must not exceed the market price cap (MPC, currently set at around $13,800/MWh) or fall below the market price floor (MPF, currently set at -$1,000/MWh).

It would be appropriate to set a floor on the local prices that arise in the OFA model. The floor could be set at the same level as the market price floor or at a different level. This issue is considered further in section 4.3.8.

Incomplete or Inconsistent Dispatch Data

Access settlements relies on data from the dispatch process. It requires that this data is complete, accurate and consistent with regional settlement.

In particular, there must be consistency between generation dispatch, the RRPs used in regional settlement and the flowgate prices used in access settlement. As explained in section 4.2.2, this ensures that the "no-regret dispatch" principle is maintained; that generators are paid the LMP at the margin for their output.

There are a number of situations where dispatch data may not be accurate, complete and consistent. These situations are summarised and categorised in the table below.

Figure 4.8 Incomplete or inconsistent dispatch data

	Incomplete	Re-run	Admin	OCD
Manifestly Incorrect Inputs	✓			
Scheduling Error	✓			
Market Suspension	✓			
Administered Price Cap			✓	
MPC Over-ride			✓	
Price Scaling (across interconnector)			✓	
Intervention Pricing ("what-if")		✓		
Over-constrained Dispatch (no re-run)				✓
Over-constrained Dispatch (after re-run)		✓		
RRP cap or floor binds			✓	

The four categories have the following meanings:

- *incomplete:* there is not a complete and consistent database of the dispatch and metering data required for access settlement;
- *re-run:* dispatch prices and RRP are based on a re-run of NEMDE, whereas dispatch targets, and so generation output, are based on the original dispatch run;

- *admin:* RRP is set to an administered level which differs from the LMP at the RRN; and

- *OCD*: over-constrained dispatch means that some NEMDE constraints are violated, potentially causing extreme flowgate prices.

In these situations, whilst it may not be possible to maintain the principal of no-regret dispatch, changes to access settlement processing or algebra may mitigate the impact. Some possible changes are discussed in section 4.3.8.

4.3 Design Issues and Options

4.3.1 Flowgate Settlement

Flowgate vs Regional Balancing

Settlement balancing is a fundamental principle of OFA. The reasons for it have been discussed in section 2.3.4. Settlement balancing essentially requires that the payments *out of* access settlements, associated with flowgate entitlements, balance the payments *into* access settlements, associated with flowgate usage.

Access settlement has been designed to achieve this separately on each congested flowgate, by scaling entitlements to match usage. An alternative and, at face value, simpler approach, would be to scale access on a regional or even NEM-wide basis so as to achieve the balance.[57]

Under such an approach, there would be no need to consider flowgates at all within access settlement. Payments could simply be based on local prices and on some scaled level of access, eg:

$$Pay\$ = (k \times A - G) \times (RRP - LMP)$$

Where:

A = target access

k = global scaling factor across a region or across the NEM

Such an approach appears *mathematically* simpler than accruing settlement payments across multiple congested flowgates.

Complexity of Regional Balancing

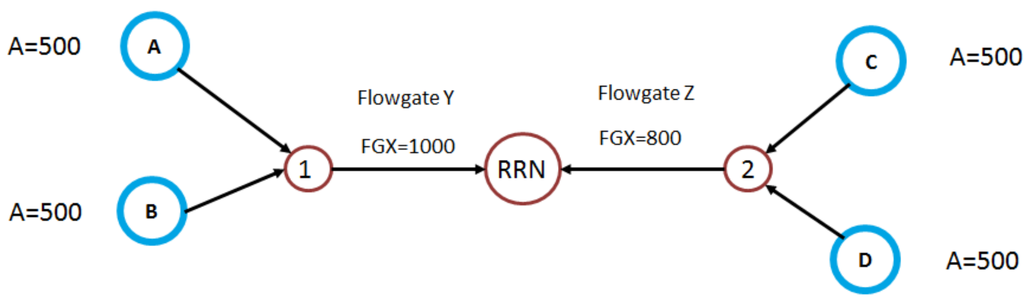
In fact, regional balancing is far more complex than flowgate balancing: not mathematically but *operationally*. An example will help to illustrate this.

57 In fact, such an approach is considered for access settlement during administered price periods and is described in section 4.3.8.

Consider Figure 4.9 below. There are two congested flowgates: flowgate Y between node 1 and the RRN; and flowgate Z between node 2 and the RRN.

There are four firm generators, as shown; some non-firm generators may also be present that cause the flowgates to be congested, but these are omitted from the diagram for simplicity. The capacity of flowgate Y is sufficient to accommodate the firm generators connected to node 1 and using that flowgate. However, the capacity of flowgate Z is insufficient to provide firm access levels to firm generators connected at node 2.

Figure 4.9 Three-node radial network



Balancing at a flowgate level allows for access levels to be maintained for firm generators at node 1 but scaled back for those connected to node 2, meaning that only firm generators C and D are impacted by the capacity shortfall on flowgate Z. A regional balancing approach would have required effective access levels of all firm generators to be scaled back by a common amount.

It is seen that the flowgate approach is not only fairer but also more efficient and transparent. It is more efficient because, in deciding on a location and a firm access level, generators will only take account of flowgate capacity and firmness on flowgates in which they participate. Access decisions should not be – and will not be – affected by congestion in other parts of the region.[58]

It is more *transparent* because access settlement occurs only on congested flowgates which, in any particular settlement period, are typically few in number. Only a subset of the congested flowgates will affect a *particular* generator. It is relatively straightforward for a generator to monitor and verify entitlement scaling and flowgate pricing on a few, relevant, flowgates. In contrast, under regional scaling, a generator would have to monitor *every* flowgate in the region.

Essentially, the problem with the "simpler" settlement formula presented above is that there is a lot of underlying complexity that is distilled into the "k" factor. Generators would need to analyse k and to do so would need to "unpack" the k-factor, analysing the flowgates, shortfalls and congestion that drive it.

58 For example, a South West Queensland generator should not have to take account of possible congestion in North Queensland.

Flowgate Monitoring

It is, in any case, naïve to think that flowgates could be ignored by generators if settlement were based only on local prices. These local prices are driven by flowgate prices and so for generators to analyse local prices and develop trading strategies and tools, they would have to analyse the flowgates that they participate in. Indeed, even under current arrangements, generators analyse the constraints and congestion which could cause them to be constrained off. Under OFA, they will necessarily be monitoring and analysing those same flowgates.

4.3.2 Defining Flowgates

Transmission Constraints

Flowgates are potential points of congestion in dispatch and correspond to the constraint equations used in NEMDE. NEMDE employs a lot of constraints and only *transmission constraints* are to be treated as flowgates.

Transmission constraints can be defined as follows:

> "any constraint that arises as a result of limitations on TNSP networks, or DNSP networks to the extent they involve dual function assets, and for which a constrained generator is not compensated under current arrangements."

Some aspects of this definition are considered in turn.[59]

First, the reference to TNSP or DNSP networks is because *un*regulated networks – ie, those operated by MNSPs – are outside the scope of OFA and any constraints arising on these are not flowgates.

Second, as explained in section 2.2.5 , *dual function assets* are referred to because these operate in parallel to transmission networks. Constraints associated with dual function assets can have a similar impact to TNSP-related constraints, are included within NEMDE and can constrain generation dispatch. Therefore, these are regarded as flowgates.

Constraints that arise on DNSP assets other than dual function assets are *not* flowgates. Although they can cause embedded generators to be constrained, OFA is not designed to provide firm access on distribution networks. These constraints are not usually included within NEMDE but are managed by the DNSP directly.

Finally, where a generator is already compensated for being constrained-off, for example in relation to NSCAS provision, the OFA model should avoid duplicating this compensation. Therefore, the associated constraints are *not* flowgates.

59 A more detailed application of this definition to existing NEMDE constraint categories is presented in appendix B.2.

Separation Constraints

FCAS constraints (those relating to frequency standards) are not generally caused by limitations on TNSP networks, meaning that they are not considered flowgates. Many FCAS constraints are affected by Basslink limitations, but Basslink is a MNSP rather than a TNSP and so, again, these constraints are not regarded as flowgates.

One type of FCAS constraint that is relevant to the OFA model is the *separation constraint*. Such a constraint may be included in NEMDE in situations where a credible contingency can lead to islanding.[60] A separation constraint sets a limit on the pre-contingent flow on the relevant network element to ensure that, should it fail, the FCAS in the two post-contingent islands can contain frequency deviations in accordance with NEM operating standards.

For the purposes of the OFA model, a separation constraint is similar to a thermal network constraint that limits the pre-contingent element flow to a specified maximum.

The difference in NEMDE is that the separation constraint is co-optimised, meaning that NEMDE can decide to source extra FCAS in order to increase the flow limit. This co-optimisation is not relevant to access settlement, which calculates the effective flowgate capacity as usual[61] and applies the access settlement algebra accordingly.

Constraint Formulation

NEMDE constraints are not necessarily in the simple form:

Total usage < RHS

For example, in the case of a separation, FCAS variables will also appear on the LHS. Also, the RHS quantity may be expressed as a complex formula, depending upon many power system variables. In a feedback constraint formulation, the RHS will include metered values of generation and line and interconnector flows.

Recall from section 4.2.5 that the RHS of constraint equations is not used in access settlement. Rather, effective flowgate capacity is calculated as the total usage of flowgate access generators and DICs. This can be derived from the metered generation and interconnector flows, and the participation factors. The latter are *always* present in constraint equations, no matter how they are formulated, because they are critical to NEMDE.

Therefore, it is not anticipated that unusual constraint formulations will create any difficulties for access settlement.

60 The splitting of the NEM into two or more separated networks.

61 Ie, based on aggregate usage from generator *energy* dispatch terms. FCAS terms are not included in the calculation.

4.3.3 Flowgate Support

Introduction

Section 2.3 discussed various similarities and differences between OFA and the Standard Market Design (SMD) common in North America. One difference mentioned, but not discussed there, is flowgate support. This is discussed below.

Local Prices

SMD is a nodal market for generation: all generators are paid their LMP for their output. On the other hand, OFA pays LMP to some generators only:

- flowgate access generators (LMP<RRP) are paid LMP; and
- flowgate support generators (LMP>RRP) are paid RRP.

Put another way, all generators are paid the lower of RRP and LMP.[62]

There are several reasons why it has been decided to design OFA differently to the SMD in this respect:

- concern about the pricing power of constrained-on generators;
- transition issues;
- use of flowgate support agreements as an alternative; and
- difficulty in designing a standard firm support service that would be attractive to generators and TNSPs.

These reasons are discussed in turn below.

Pricing Power

A generator that is in a *load pocket*, a demand-rich area with limited transmission capability, may frequently be constrained on in order to maintain local reliability. Such a generator may have substantial, possibly extreme, *local pricing power*[63] and, were it paid the LMP, might profitably use this power to raise its local price.

It is common in SMD markets for this pricing power to be regulated as part of the market design. For example, caps may be placed on the local price or offer price of

62 This is something of a simplification, since under conditions of mixed congestion, a generator may be simultaneously a flowgate access generator on one congested flowgate and a flowgate support generator on another. Mixed congestion is discussed further in appendix A.1.

63 Pricing power means ability to change the market price by varying its offer. Local pricing power means ability to change the LMP in this way.

identified "must-run" generators, based on an analysis of their operating costs. Also, as mentioned previously, the market price cap is typically lower in SMD markets anyway.

A philosophy underlying the NEM design is to avoid regulating generating behaviour or payments, except in specific and infrequent circumstances: eg, under AEMO *directions*. It is considered that competitive discipline or (if required) the competition regulator are best for managing behaviour.

Transition

Currently, flowgate support generators are paid RRP.[64] Paying them a higher price under OFA would provide them with a windfall gain. That could be potentially managed through transitional arrangements.[65]

Flowgate Support Agreements

In the NEM currently, TNSPs enter into Network Support Agreements (NSAs) with generators in load pockets where the TNSP needs them to run occasionally so as to maintain network reliability standards. It is anticipated that, under OFA, TNSPs could extend this practice, entering into new *Flowgate Support Agreements* with flowgate support generators whose output is required in order to maintain the FAS.

This approach – of *ad-hoc, negotiated* agreements between generator and TNSP – is the antithesis of the approach for flowgate access generators, for whom a *standardised, regulated* firm access service is established. So why is a different approach appropriate for flowgate support? There are several reasons.

First, as a matter of history, generators have successfully negotiated network support agreements with TNSPs. However, no equivalent network access agreements have been entered into on this basis. So it is reasonable to expect that flowgate support, unlike flowgate access, can be managed effectively without the need for detailed regulation.

Second, the topology of the NEM network is such that the amount of constrained-on generation is low compared to the amount of constrained-off generation. For historical reasons, State networks are broadly hub-and-spoke, with the RRN and most of the load close to the hub. Only in Queensland are there substantial load pockets remote from the RRN where generation are liable to be constrained on.

Third, a regulated firm support service, analogous to the firm access service established by OFA, would be likely to be unattractive to generators or to TNSPs, as discussed next.

---

64 Except when directed.

65 As the windfall losses to flowgate access generators are managed, as discussed in chapter 9. However, this would create a new area of complexity and contention for the OFA design.

Regulated Firm Support Service

OFA *could* specify and regulate a firm support service that could operate in tandem with a full nodal approach. All flowgate support generators, firm and non-firm, would be paid LMP for their output.

The regulated firm support service would be provided by firm support generators to TNSPs, the reverse of the firm access service. Firm support generators who elected to provide the service would be paid a regulated fee by the TNSP. In return, these generators would be required to pay the difference between LMP and RRP on the registered support amount.

So long as it was available and dispatched, a firm a firm support generator would be paid, net, RRP: LMP on its generation minus (LMP-RRP) paid to the TNSP. But if the firm support generator were unavailable, it would still be liable make payments to the TNSP. The generator would find itself short against, possibly extreme, LMP.[66]

Designing firm support agreements to be inactive when the generator is unavailable would avoid this short risk and make the service more attractive to generators. But such as service would be useless for the TNSP, since a generator could simply declare unavailable whenever it is constrained on.[67]

Complex mechanisms would be needed to appropriately share the availability risks between TNSP and generator. Such mechanisms are best developed through a negotiation (of a flowgate support agreement) rather than through regulation.

4.3.4 Counter-price Flows

Introduction

A *counter-price* flow on an interconnector means a flow directed from the high price region to the low price region, the reverse of the usual, and intuitive, direction. Since the IRSR is the product of the interconnector flow and the inter-regional price difference, counter-price flows usually lead to a *negative* IRSR: ie, a *deficit* in regional settlement.

Dispatch Efficiency

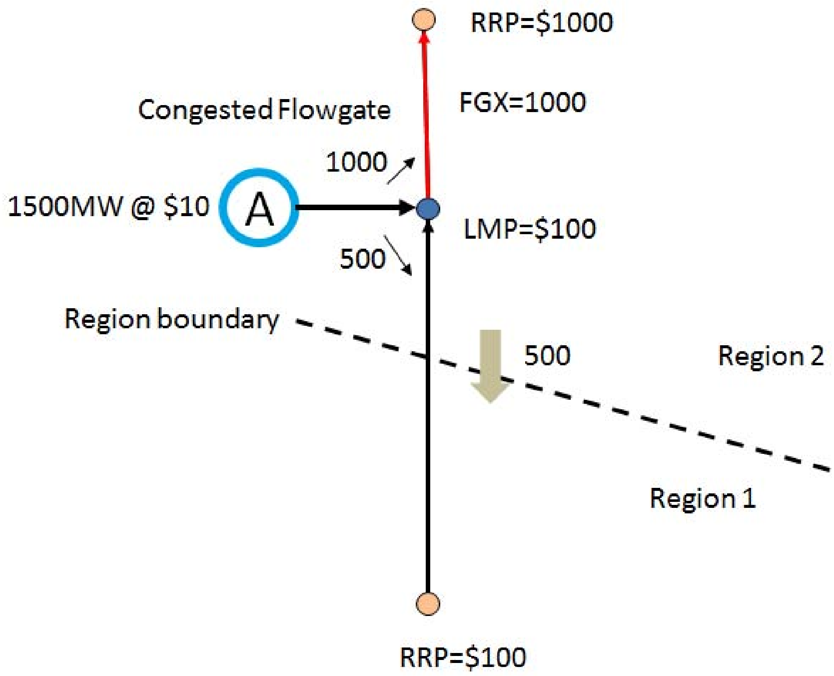
Counter-price flows tend to be associated with dispatch inefficiency. It certainly *appears* to be inefficient to transfer power from a high price to a low price region. However, regional prices are unlikely to be representative in this case. Typically, a counter-price flow is caused by a low-price generator behind an intra-regional constraint being

66 This is analogous to the situation where a generator that has sold hedges to retailers can find itself short against RRP if it is not available and so is unable to back those hedges.

67 As generators do currently.

dispatched and so forcing a reversal of the interconnector flow, as illustrated in Figure 4.10.

Figure 4.10 Counter-price flow



In this figure, generator A is located on the interconnector path and has an offer price ($10/MWh) lower than both RRPs and so is fully dispatched. Because of a congested flowgate to the north of the generator, part of its output flow south, across the regional boundary. This appears as a counter-price flow.

Under current arrangements, a generator may bid below cost to maximise its dispatch and hence access. Generator A in the example *might* have a cost above $100/MWh, but bids below this level anyway in order to maximise its dispatch by creating a counter-price flow. So, under current arrangements, the dispatch *might* be inefficient during counter-price flows.

By separating access from dispatch, OFA removes this incentive to bid low. Counter-price flows may still occur under OFA, but only if the constrained local generator is *actually* cheaper than the interconnector. In this case, despite the counter-price flow, dispatch is *efficient.*

Under OFA, the impact of a counter-price flow depends upon the EFGX of the congested inter-regional flowgate: specifically, whether the EFGX is positive or negative. These two situations are considered in turn below.

Positive Flowgate Capacity

The payment to a DIC on a congested inter-regional flowgate equals the product of the entitlement and the flowgate price. On a congested flowgate, the flowgate price is always positive. When there is positive EFGX, the DIC entitlement is non-negative, as discussed in section 4.2.3. So, despite the negative IRSR, DIC payments are non-negative.

This is possible because, as well as receiving the (negative) IRSR, the DIC receives a payment from access settlement on the flowgate. This payment would be funded by another participant on the flowgate. Typically, this would be a generator that has been dispatched above its entitlement: because it is non-firm or because a flowgate shortfall means that its entitlements has been scaled back.

In the example above, flowgate capacity on the congested flowgate is positive. Generator A would have an entitlement on that flowgate of at most 1000MW and has usage of 1,500MW. Therefore, it will pay *into* access settlement, offsetting the negative IRSR.

Negative Flowgate Capacity

As discussed in section 4.2.4, when there is negative flowgate capacity, the interconnector must provide flowgate support and so the relevant DIC (who is directed in the direction of the interconnector flow) receives nothing from flowgate settlement: it is simply allocated the negative IRSR. The TNSP in the importing region (ie, the region towards which the interconnector is flowing) is responsible for funding this negative IRSR.[68] This is, in practice, the same outcome as under the current arrangements. However, the rationale is rather different and is explained below.

An intra-regional analogy will help to illustrate this rationale. Intra-regionally, negative flowgate capacity will lead to a generator being dispatched to provide flowgate support. Without this support, the local load would have to be shed, so the generator is in fact supporting transmission reliability. In these circumstances, the TNSP would typically enter into a network support agreement (NSA) with the generator: to ensure it is available and to compensate it for being constrained-on.[69]

Unlike a generator, a constrained-on DIC[70] cannot declare itself unavailable in order to avoid the cost of the negative IRSR. But, aside from this difference, there is a strong analogy between the intra-regional and inter-regional cases. The DIC is constrained on to support reliability in the importing region, which is the responsibility of the

---

68 Under revenue regulation, the TNSP would be permitted to pass-through the cost of the negative IRSR to TUOS customers.

69 In the absence of the NSA, a constrained-on generator would simply declare itself unavailable and would have its output directed by AEMO. The generator would be entitled to compensation for this.

70 The DIC is constrained on in the sense that it is dispatched when the RRP it is paid (the RRP in the importing region) is lower than its “cost” (the RRP in the exporting region).

importing TNSP. There is no network support agreement but, by funding the IRSR, the TNSP is essentially paying the DIC for its network support.

Clamping

Under current arrangements, AEMO is required to *clamp* the interconnector when this is necessary to prevent the negative IRSR becoming too large. Clamping is done by introducing an artificial constraint into NEMDE which prevents or limits the counter-price flow. The objective of clamping is to prevent undue risk being placed on TNSPs and on their TUOS customers, to whom the cost is passed-through. In the case where the counter-price flow is symptomatic of inefficient dispatch – as discussed above - this clamping may improve dispatch efficiency, although this is not its intent.

Under OFA, there is similarly a concern about the financial risks associated with the negative IRSR. However, these only arise in the case of negative flowgate capacity, for reasons discussed above. In these circumstances, clamping is inadvisable, since it will simply lead to load shedding: the flowgate support from the DIC is essential to maintaining transmission reliability.

Therefore, the financial risks need to be managed another way. This might be through the TNSP:

- strengthening the network to avoid negative flowgate capacity;
- procuring demand-side management in the locality to improve flowgate capacity; or
- entering into financial hedges with generators in the exporting region to hedge the IRSR volatility.

It would be a matter for the TNSP to select and undertake the appropriate action.

In summary, clamping should not be used in OFA. TNSPs should find other ways to manage their market risks during periods of negative flowgate capacity.[71]

4.3.5 Target Access

Overview

As was discussed in section 2.3.2, despite commonly being referred to as a compensation mechanism for constrained-off generators, access settlement is not explicitly designed in this way. Rather, it makes payments to firm generators that are similar to those that would be made to FTR holders under the SMD.

71 Indeed, since interconnector clamping is not effective when there is negative flowgate capacity, TNSPs are already bearing and managing these risks currently.

This section discusses whether access settlement could, or should, be designed to be more faithful to a compensation model, by changing the basis for setting target firm access. It also discusses the setting of target non-firm access.

Three possible alternative measures are considered for capping access:

- preferred output;
- availability; and
- uncapped.

These are discussed in turn below.

Preferred output

To know whether a generator is constrained off, one must estimate its preferred output: the level of dispatch that the generator would choose if it were not constrained. Taking its offer at face value, the preferred output is the amount of capacity that is offered at prices below the current RRP.[72]

To design access settlement as a compensation mechanism, the target access of a firm generator should be capped by the preferred output, so that it is compensated when it is constrained off but not otherwise: eg, when it is out-of-merit or unavailable.

The practical difficulty with estimating preferred output using the dispatch offer is that this is easily manipulated by rebidding. An available, but out-of-merit, generator could simply rebid[73] to raise its preferred output to equal its availability. In practice, all that is being measured is the generator availability. This measure also prompts rebidding which could become disorderly: eg, if the congestion is only ephemeral.

An alternative approach would be to estimate preferred output based on an estimate of generator costs. However, undertaking such cost estimation would be intrusive and complex and is not in keeping with the NEM design of a bid-based rather than a cost-based market.

To conclude, it is not feasible or appropriate, in a bid-based market, to design access settlement as a compensation mechanism for constrained-off generators based on generators' preferred output.

72 For simplicity, dynamic constraints on generation, such as minimum stable generation and ramp rate limits, are ignored in this discussion.

73 It can do this by bidding at a price anywhere between LMP and RRP so that it appears to be constrained off.

Availability

Alternatively, target access for firm generators could be based on availability – as it is for non-firm generators. This is a better approximation to preferred output than is generator capacity, since at least unavailable generators will not be "compensated".

This would be the preferred approach if "compensation" actually were the main objective of access settlement. But, the OFA philosophy has progressively moved away from "compensation" towards something more akin to the FTRs in the SMD.

If access settlement for firm generators is based on availability, then the value of firm access to a generator, and the cost to a TNSP of providing it, is now predicated on the reliability of the generator. Should the access price then reflect the generator's reliability? What happens if firm access is purchased by an unreliable generator (at a relatively low price) and then sold to a reliable generator?

These difficulties are not insurmountable, but beg the question as to whether "compensation" is the right objective. An FTR-type approach simplifies access settlement, access pricing and trading. It is not clear that it would substantially degrade the value or firmness of the firm access service. As discussed in section 2.3, SMD is a common and successful market model. There need to be special circumstances in the Australian context to move away from it. It is not clear what these would be, in this instance.

Uncapped

OFA does, nonetheless, depart from the SMD in one aspect of target access. In OFA, target access is capped by generator capacity; there is no corresponding limitation on FTR payouts (or purchases) under SMD.

The capacity-based cap on target access was originally introduced to the OFA model as the best practical approximation to preferred output, given that other approaches having been rejected for the reasons discussed above. The OFA is now so far removed from the compensation model that removing the capacity-based cap could be seen as the logical next step.

However, there is some concern about the ramifications of this. The capacity-based cap effectively places some restrictions on the procurement and trading behaviour of generator[74] and such restrictions may be prudent when such behaviour is uncertain. Possibly, the capacity restriction could be removed at a later date.

74 There is no explicit restriction, but a generator may be reluctant to purchase firm access in excess of their capacity when they get no benefit from this in access settlements.

Target Non-firm intra-regional Access

It is proposed that target non-firm access is based on *availability*. The difficulties with using such an approach for target *firm* access, discussed above, do not apply to *non*-firm access, since non-firm access is neither purchased nor traded.

Some reasons favouring the proposed approach are:

- it is the best available proxy for preferred output, which is a *fair* basis for allocating non-firm access; and
- it is similar to the *de facto* allocation of access under the current arrangements during disorderly bidding, assuming that generators have identical participation factors.

It is acknowledged that *fairness* is not specifically an objective of NEM design, but in the context of non-firm access – where there are unlikely to be any significant efficiency implications – it seems a reasonable criterion. It is for this reason that it is proposed to retain the availability-based approach to target non-firm access. However, this is not a critical design element and a capacity-based cap could be adopted instead.

Zero target non-firm access for interconnectors

The OFA design generally aims to minimise the difference in treatment between generators and interconnectors, in order to avoid discriminating for or against inter-regional trading. In the light of this, the different treatment of interconnectors in relation to target non-firm access needs some explanation.

Firstly, target non-firm access for generators is based on availability, but there is no obvious measure of availability for interconnectors. In a sense, the availability of an interconnector is infinite, since its flow can only be constrained by congestion on inter-regional flowgates. Certainly, interconnectors do not *offer* availability like generators do, so there is no ready measure for use in access settlements.

Secondly, the component of the payment to DICs that relates to non-firm access is passed onto TNSPs who, in turn, pass it on to TUOS customers, who gain no hedging value from it. In contrast, payments to non-firm generators do have some hedging value. Therefore, it makes sense to maximise the payments to non-firm generators, at the expense of DICs.

The current SRA arrangements reflect this preference. Although allocated initially to TNSPs, the IRSR is then sold through the SRA, so that any hedging value that it provides is made available to market participants, rather than being lost in TUOS.

The proposed approach therefore appropriately maximises the value of non-firm access, by setting the target non-firm interconnector access to zero and maximising the payments made to non-firm generators.

4.3.6 Super-firm Access

Overview

Target firm access is capped by generator capacity. So, although a generator could decide to be *super-firm* by procuring a firm access higher than its capacity, this additional amount attracts no payments from access settlement payments.

An alternative approach is for access settlement to recognise this higher purchase by providing a *super-firm access* service to super-firm generators. Under this service, the maximum access provided is still limited by generator capacity, but access would not be scaled back to the same extent during flowgate shortfalls.

This section discusses how such a super-firm service would be provided through access settlements and the implications for OFA design and outcomes.

The super-firm issue is only relevant to intra-regional firm access. DICs do not have any capacity limit and so cannot be super-firm, by definition.

Super-firm Entitlements

Access settlement for a super-firm service would include and calculate a *super-firm entitlement* for super-firm generators. This would be added to the firm and non-firm entitlements allocated for that generator. When allocated, these super-firm entitlements would increase access settlement payments to the generator, accordingly.

The process for determining these super-firm entitlements is similar to the usual entitlement allocation process:

1. A *target super-firm access* amount is defined, being the amount (if any) by which registered access exceeds generation capacity.

2. A *target super-firm entitlement* is defined on a congested flowgate, being the target super-firm access amount multiplied by the generator's participation factor.

3. The actual super-firm entitlement is no higher than the target and is scaled back as needed when there is a flowgate shortfall. The same scaling factor is applied to super-firm and firm entitlements.

4. The actual super-firm entitlement is further reduced, if needed, to ensure that the access provided to the generator does not exceed it capacity.

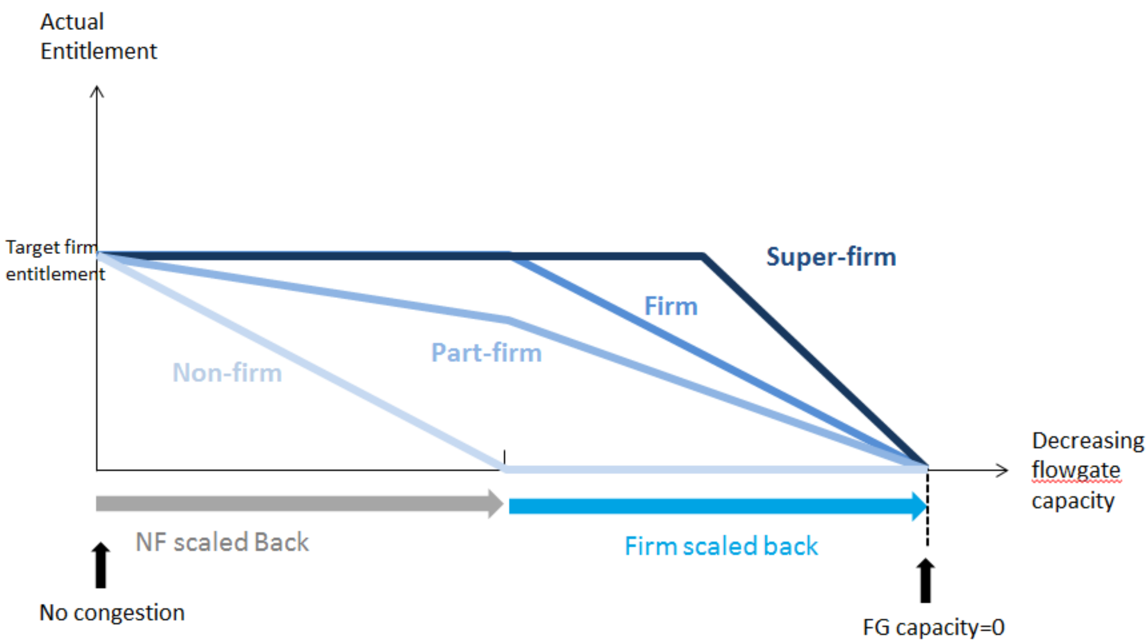
This algorithm can be illustrated by a simple example. A 100MW generator on a radial constraint chooses to be super-firm, purchasing 120MW of firm access and so enjoys a super-firm service. Its firm and super-firm access targets are 100MW and 20MW, respectively. Since it is a radial constraint (unity participation factor) its target entitlements are the same.

If there is no shortfall on the flowgate, the generator is allocated its full firm entitlement of 100MW. It is not provided any super-firm entitlement, since doing so would cause its access to exceed its capacity.

On the other hand, suppose that there is a shortfall on the flowgate, and that firm and super-firm access is scaled back by 50% as a result. In this case, it receives 50MW of firm entitlement and 10MW of super-firm entitlement. In aggregate, it receives 60MW of entitlements. So, it receives a higher level of access than an equivalent firm generator, who would have received only 50MW in these circumstances. In general, super-firm generators will not have their access scaled back as severely during shortfalls as firm generators, and so they genuinely receive a super-firm service.

Figure 4.11 presents the super-firm service graphically, comparing the entitlement scaling for a super-firm generator with that for other generator categories

Figure 4.11 Entitlement scaling four categories



Detailed algebra for allocating super-firm entitlements is presented in appendix A.3.

Implications of Super-firm

Allocated entitlements must, in aggregate, equal EFGX. If one generator is receiving an additional super-firm entitlement, other generators must be allocated lower entitlements as a consequence. This would occur through a reduction in the firm scaling factor.

Therefore, to avoid super-firm access causing the firm access service to be less firm, the super-firm purchase must be recognised in the firm access standard. The TNSP should be required to provide additional flowgate capacity in order to provide the super-firm service. This is discussed in section 5.3.3.

In recognition of this, a super-firm generator should be charged more for access than a firm generator. In fact, the access pricing method ignores generation capacity, so a 100MW generator buying 120MW of access will face the same access price as a 120MW generator buying the same amount. This is discussed further in chapter 6.

Assessment of Super-firm

Super-firm access gives generators the *option* of having access that is *firmer* than the FAS provides. The benefit of this feature depends upon how firm the FAS actually is, compared with how firm a service generators desire. FAS is described in the next chapter; the preferences of generators are, of course, uncertain.

Super-firm access creates some additional complexities in the OFA design: in access settlement, as described here, and in FAS. Given that these complexities are significant, and the benefits of super-firm uncertain, it has been decided not to include super-firm in the OFA design. Potentially, if there were a clear demand from generators for a super-firm access product, super-firm could be introduced into OFA at a later time.

4.3.7 Thirty Minute Settlement

Overview

Access settlement relies on information from dispatch and metering. Some of this information is available for each dispatch interval (DI – 5-minute period) and some only for each trading interval (TI – 30-minute period), as presented in Table 4.6 below.

Table 4.6 Information available for settlement

Variable	Period Available
Sent-out generation	TI
Availability	DI
Participation factor	DI
Flowgate price	DI

It is proposed that access settlement uses a settlement period of a TI. This is consistent with regional settlement, which is also settled by TI, although the FCAS market is settled by DI. Dispatch variables that are sourced on a DI-basis must be converted to a TI-basis before being applied to access settlement. It is proposed that this is done using *unweighted* averaging: eg, the FGP for the TI is simply the arithmetic mean of the FGPs in the six DIs within the TI.[75]

75 Where a value for FGP does not exist for a flowgate in a DI - because the relevant constraint is not invoked - its value is taken to be zero for the purposes of calculating the TI values.

There are two other possible approaches to access settlement which are considered in this section:

- weighted-TI: on a TI-basis, using averages that are weighted by FGP; or
- DI: on a DI basis.

These three possible approaches are summarised in Table 4.7 below.

Table 4.7 Three possible approaches to access settlement

Variable	Source basis	Unweighted TI	Weighted TI	DI
G	TI	Source	Wtd mean of deemed DI values	Deemed DI values
Availability	DI	Unweighted mean	Weighted mean	Source
Participation	DI	Unweighted mean	Weighted mean	Source
FGP	DI	Unweighted mean	Unweighted mean	Source

In two of the options, deemed-DI values are required for generation. These values would be determined such that:

- in aggregate, they sum to the TI source value; and
- in profile, they follow a DI-based generation measure.

The DI-based generation measures could be based on either SCADA metering values or dispatch targets, both of which are available on a DI basis.

In the weighted-TI approach, FGP is used as the weighting factor.

The weighted-TI approach will give similar – but not identical – outcomes to the DI approach. It can, perhaps be thought of as "DI-lite": approximating to DI-settlement but appearing to be like TI-settlement. The differences are detailed in appendix B.3.

Flowgate Congestion and Capacity

Recall that, in access settlements, EFGX is calculated based on usage rather than based on the RHS of the constraint equation, ie:

EFGX = total usage of flowgate access generators

Total usage equals flowgate capacity only when the flowgate is congested. When the flowgate is uncongested, total usage is below capacity: by definition. On the other hand, in a DI in which the flowgate is not actually *invoked* in NEMDE[76], total usage can be higher or lower than flowgate capacity.

Since access settlement takes place on congested flowgates, TI-settlement must define what exactly this means. It is proposed that a flowgate is considered congested for a TI if it is *congested* in at least one DI within the TI. In the remaining DIs the flowgate could be:

- also congested: total usage equals flowgate capacity;
- uncongested: so total usage is below flowgate capacity; or
- not invoked:[77] so total usage could be higher or lower than flowgate capacity.

On average, over the TI, total usage could be higher or lower than FGX. If it is lower, this may mean that access is scaled back more than is really necessary.

This error in estimating FGX could be avoided by using the *true* flowgate capacity (ie, the constraint RHS) when allocating entitlements. But then aggregate entitlements would no longer match aggregate usage and so access settlements would not balance. This would be an unacceptable breach of the settlement balancing principle.

In summary, the unweighted-TI method necessarily leads to errors in estimating of flowgate capacity. The materiality of these errors has not been assessed.

This problem does not arise under DI-settlement: flowgates are only settled in the DIs in which they are congested. The problem is also addressed under weighted-TI. Since FGP is zero in the DIs for which the flowgate is uncongested or not invoked, the EFGX in these DIs has no impact on the *weighted*-average.

Basis Risk

OFA provides a generator with access to the RRP. RRP is defined, on a TI basis, as the *unweighted* average of *dispatch prices* from each DI within the TI: the dispatch price is the LMP at the RRN in a particular DI. Recall that a generator with network access is paid an amount according to the formula A × (RRP – LMP). Under unweighted-TI settlement, the RRP in this formula is also an unweighted value. But, under DI or weighted-TI settlement, the RRP is instead based on a weighted-average of dispatch prices. The weighting factor will be based on volume measures such as generation output and EFGX.

Therefore, under these alternative two settlement approaches, a generator is exposed to some *basis risk:* between the unweighted-RRP against which forward contracts are struck, and the weighted-RRP which arises under access settlement. The risk arises when there is a significant difference between the weighted and unweighted RRPs.

76 That is to say the constraint exists in the constraint library but is not applied to NEMDE in the particular DI.

77 Eg, because the constraint is only invoked for the last DI in the TI.

This will occur when dispatch prices are volatile and the volume measures are variable, with a TI. The materiality of this risk has not been assessed.

Generation Data

Access settlement is based on the TI-based revenue metering because this is currently used for regional settlement and existing mechanisms ensure that it is reliable and accurate. DI and weighted-TI access settlement would also need to source DI-based generation data: either SCADA metering or dispatch targets. This data may be less reliable than revenue metering and may introduce errors or uncertainties into access settlement.

Conclusion

It is proposed to use the unweighted-TI approach for access settlement. This is on the basis of its greater simplicity, avoidance of basis risk, and use of reliable generation data. There are some concerns about the errors in EFGX calculation associated with this method. Further analysis could assess the relative materiality of the basis risk and EFGX errors and this might prompt a different design decision.

4.3.8 Abnormal Settlements

Overview

As described in section 4.2.5 under *abnormal conditions* there will be a lack of consistency or completeness in the data provided from the dispatch and regional settlement processes that are used by access settlement. Four different types of abnormality can arise:

- *incomplete*: there is not a complete and consistent database of the dispatch and metering data required for access settlement;
- *re-run:* local prices, dispatch prices and RRP are based on a *re-run* of NEMDE, whereas dispatch targets, and so generation output, are based on the original run;
- *admin:* RRP is set to an administered level which differs from the LMP at the RRN; and
- *OCD:* over-constrained dispatch means that some NEMDE constraints are violated, potentially causing extreme flowgate prices.

Furthermore, as discussed in section 4.2.5, even when the data is consistent and complete, there can be extreme flowgate prices and local prices. It may be appropriate to cap, or floor, these prices in access settlements.

Some possible changes to access settlements to mitigate these issues are considered for each abnormality type in turn.

Extreme Prices

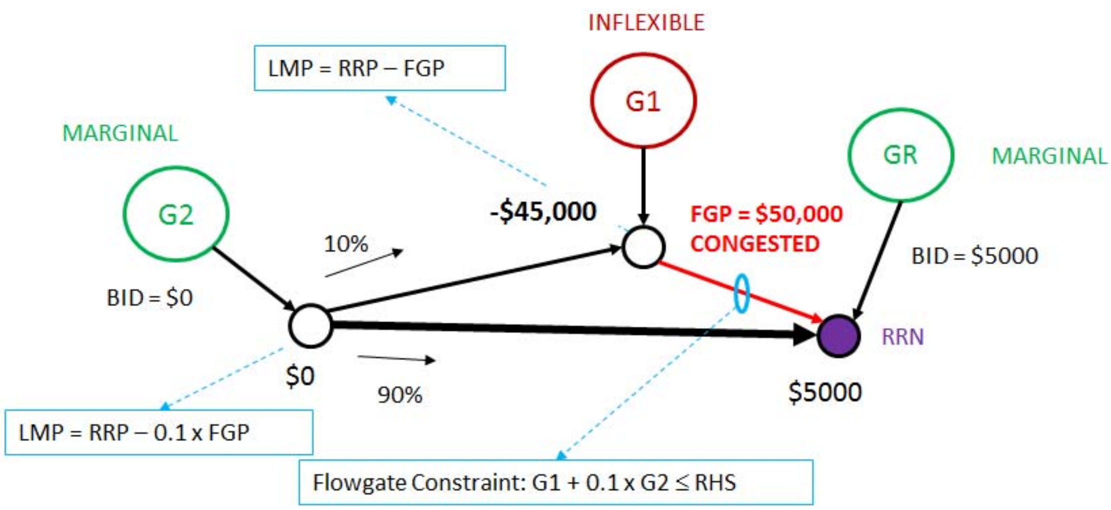
Under OFA, flowgate access generators are paid, at the margin, their LMP for their generation output.[78] Therefore generators, and particularly non-firm generators, are financially exposed to extremely *low* LMPs.[79]

Local prices can fall to extremely low levels through a combination of high RRPs and a *gearing* effect when the marginal generator behind a constraint has low participation. An example of this is presented in Figure 4.12 below.

In this example, a marginal generator, G2, with an offer price of zero, has only a 10% participation in the congested flowgate. This means that a 1MW increase in the flowgate capacity would allow a 10MW increase in the marginal generators output, displacing 10MW from the marginal generator, GR – offering at $5,000 – at the RRP. The flowgate price is set at the value of the associated cost saving: 10MW x $5,000 = $50,000.

The high FGP impacts directly on another generator G1, which has a unity participation in the flowgate. Each extra MW of output of G1 has the same impact as a one MW reduction in flowgate capacity (ie costing $50,000) but only saving 1MW of output from GR. Therefore the LMP for G1 is -$45,000: many times lower than the market price floor of -$1,000.

Figure 4.12 Extreme LMP



78 See section 4.2.1.

79 Recall that the treatment of flowgate support generators means that generators are paid the *lower* of RRP and LMP, so the impact of extremely high LMPs can be no worse than the impact of extremely *high* RRP currently.

Since the extreme FGP in a sense causes the problem of the extreme local price, it can form the basis for a solution. Specifically, if FGP is *reduced* then the local price *increases*. To increase an extreme local price up to a minimum *local price floor*, the FGP can be scaled *down* in accordance with the formula:

$$FGP_{revised} = (RRP - LPF)/\alpha$$

Where:

LPF is the local price floor

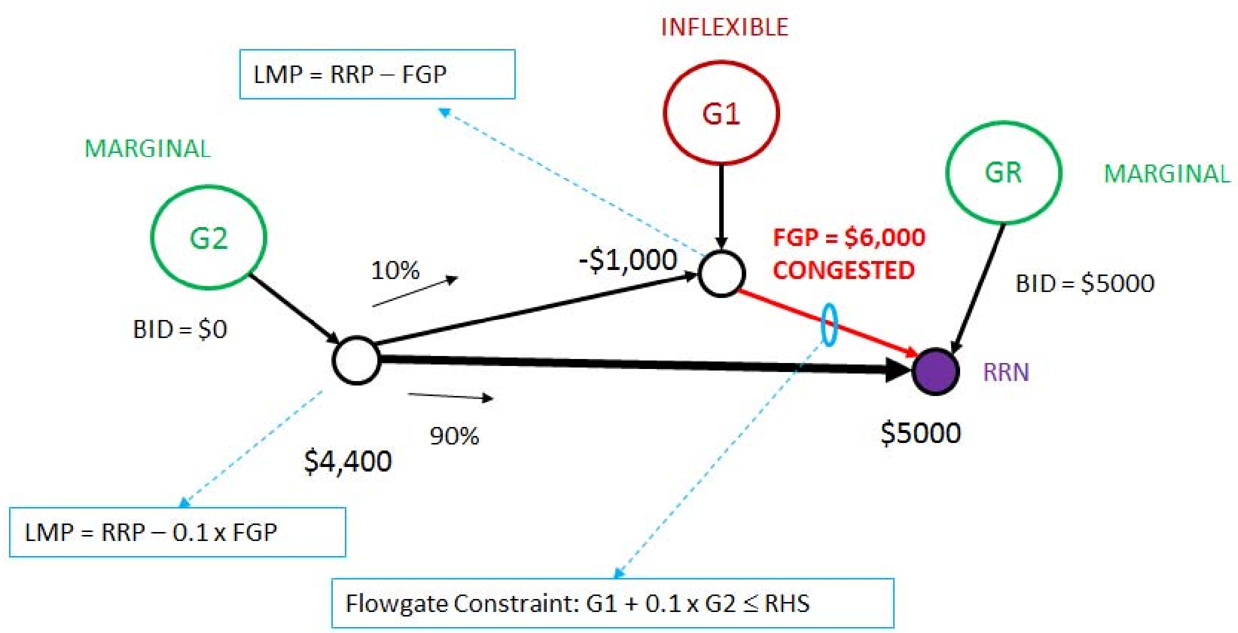
α is the participation factor of the generator that has the extreme local price

Suppose that the LPF is set equal to the MPF at -$1,000. Then, in the example, the FGP must be revised down from $50,000 to $6,000.[80] This revised FGP must apply to all access settlement payments at the flowgate, in order that access settlement continues to balance. This will cause:

- an *increase* in settlement payments to those generators whose entitlements exceeds their usages: eg, dispatched non-firm generators; and
- a *reduction* in settlement payments to those generators whose entitlements exceeds their usages: eg, non-dispatched firm generators.

In the example, as presented in Figure 4.13 the FGP scaling increases the local price for G1 from -$45,000 to -$1,000.

Figure 4.13 Floored local price



80 (RRP-LPF)/α = ($5,000+$1,000)/1=$6,000.

But it also increases the local price for G2 from $0 to $4,400.[81] If G2 is firm, it will hope to be compensated for being constrained off, by at least at the difference between RRP and its bid: ie, $5,000/MWh. However, because of the scaled FGP, it is only compensated at $600/MWh.[82]

The process for revising flowgate prices becomes more complex when the generator suffering the extremely low local price is participating in multiple congested flowgates. A generalised process is described in the appendix B.4.

Recall that the LMP is a *clearing price*, meaning that generators should not be dispatched when their offer price is higher than the LMP. In the NEM, offer prices below the MPF are not permitted.[83] Therefore, it would seem that a generator whose LMP is below the MPF should not be dispatched and so should not be exposed to the extreme LMP.

However, this analysis ignores *dynamic constraints* on generation dispatch: for example, a minimum generation constraint[84] or a ramp rate limit. These can prevent a generator that is currently dispatched from being dispatched to zero when the extreme LMP occurs. If LMP were not floored, generators would likely procure firm access to at least their minimum generation level in order to provide protection against these extreme LMPs; possibly higher if they had slow ramp rates.

A LPF could be set according to a process established in the Rules, as the MPF is currently. The process would need to balance two factors. Firstly, the risks of extreme LMPs for non-firm, inflexible generators. Secondly, the risks of FGP scaling for firm generators. In effect, FGP scaling makes firm access less firm. Furthermore, extreme LMPs are likely to occur at times of extreme high RRP: the times when access firmness is most critical.

At this stage, whilst it is clear that there should be a floor on local prices in access settlement, the appropriate level of the floor is less clear. Although setting the LPF equal to the MPF is simple and intuitive, it is not necessarily appropriate, given the different factors that apply to the LPF, as discussed above.

Incomplete Data

In the *incomplete* situations, dispatch data is incomplete or inaccurate. When the market is suspended in a region, there may be little valid dispatch data available and so access settlement must be suspended in that region. Access settlement could potentially continue in unaffected regions.

81 LMP = RRP - α x FGP = $5,000 - 0.1 x $6,000 = $4,400.

82 Compensation = α x $FGP_{revised}$ = 0.1 x $6,000 = $600/MWh.

83 After adjusting for losses.

84 Minimum generation is the lowest output that a generator can provide without going off-line.

Re-run Dispatch

In "what if" *re-run* situations, there are two dispatch runs undertaken for the same dispatch interval. The *original dispatch run* is used in determining dispatch targets. A *re-run dispatch* is used to calculate RRPs. Therefore, although there are flowgate prices available from the re-run dispatch data which are consistent with the regional prices, this entire set of prices is inconsistent with generation dispatch. This means, for example, that a generator may be dispatched even though its local price (as specified in the re-run dispatch) is below its offer price; a breach of the principle of no-regret dispatch.

Differences between the original and re-run dispatch targets are generally modest and existing mechanisms are in place to compensate generators that are materially impacted. Therefore, it is proposed that the access settlement is undertaken in the usual way, based on information from the *re-run* dispatch.[85] Existing compensation mechanisms should be extended to cover any adverse impacts on generators arising under access settlements during the re-run situation.

Administered RRP

Normally, RRP in each region is set equal to the LMP at the RRN: referred to as the *Regional Original Price* (ROP). In various situations, RRP is *administered:* set to a price other than the ROP. This creates an inconsistency between flowgate prices and RRPs. It means that, if access settlement operates as normal, generators will not be paid their LMP at the margin: their local price would differ from LMP by the amount of the difference between RRP and ROP.[86] This difference can be large. For example, under an administered price cap, the RRP could be capped at $400/MWh, say, when the ROP is as high as $13,000/MWh.

It is proposed that a change to the access settlement algebra applies when RRP is administered. The revised algebra ensures that:

- generators whose LMP is less than the (administered) RRP continue to receive their true LMP at the margin; and
- generators with LMP higher than RRP will just be paid RRP, reflecting how flowgate support generators are paid under normal conditions.[87]

On their own, these adjustments would give rise to an imbalance in access settlement. To restore balance, the effective access levels of generators must be adjusted. It is

85 Of course, output quantities will be based on meter readings, which will reflect the original dispatch: hence the inconsistency.

86 This is explained in appendix B.5.

87 Because of the administered RRP, though, it will not just be flowgate support generators in this situation. A generator with LMP less than ROP (ie, a flowgate access generator) may nevertheless have LMP greater than RRP: eg, under an administered price cap: LMP=$1,000; ROP=$10,000; RRP=$400.

proposed that this is done by using a regional scaling factor to scale effective access for every generator, and every importing FIR, by a common amount, such that access settlement balances in each region.[88] The revised algebra is presented and explained in appendix B.5.

With this revised algebra, the key OFA principles of settlement balancing and "no-regret dispatch" are maintained, except in relation to generators whose LMP is *higher* than the administered RRP. Compensation is available, under existing mechanisms, for those generators, who, in effect, become constrained-on: dispatched despite RRP being below their offer prices. It needs to be considered whether these mechanisms needed to be extended or adjusted to include any new impacts arising under the OFA model.

Over-constrained Dispatch

AEMO configures NEMDE so that it is permitted to violate some dispatch constraints if there is no *feasible* dispatch solution that complies with *all* constraints. A notional dispatch cost, referred to as a *constraint-violation penalty*, is incurred when a constraint is violated. The level of this penalty is set at a very high level – a multiple of the market price cap – so that NEMDE only violates constraints when no feasible solution exists.

If a transmission constraint is violated, the rate of the constraint violation penalty will be incorporated into the flowgate price. Using such extreme flowgate prices in access settlement would give rise to extreme and inappropriate payment amounts.

Under current arrangements, when an OCD condition occurs, AEMO re-runs dispatch, *relaxing* the problematic constraint by increasing the RHS to the extent necessary. This removes the OCD condition and removes its impact on RRP and FGPs. AEMO does this to remove the impact of OCD on RRP and also to remove the impact on FGPs: the latter is undertaken to facilitate settlement of the AER's STPIS scheme, which provides incentives on TNSPs to efficiently schedule planned outage and which depends upon FGPs.

Note that, as with the re-run situation described above, the re-run dispatch is used only for price setting; the original dispatch is used for setting dispatch targets. Thus, in this case, the OCD situation is identical to the re-run situation and the proposed approach to access settlement is the same: ie, use FGPs from the re-run dispatch.

88 The regional scaling factor can be higher or lower than one, depending upon, first, whether the RRP is higher or lower than ROP and, second, on the relative impacts of congestion on firm versus non-firm generators.

4.3.9 Loss-adjustment of Access

Overview

As discussed in section 4.2.5 both usage and target access are adjusted to reflect losses. This is done by multiplying them by each generator's Marginal Loss Factor (MLF). The adjustment to usage is needed to ensure no-regret dispatch, as discussed in appendix A.4. The adjustment to entitlements is not so critical, but is thought to be desirable. This section considers the benefits of this latter loss-adjustment. It considers, in turn, the impacts on:

- constrained-off compensation;
- access certainty; and
- FAS.

Constrained-off compensation

If target firm access is loss-adjusted, then the access level provided by firm access (through access settlements) is similarly adjusted. For example, suppose a 100MW generator has an MLF of 0.95. The loss adjustment means that it will only receive, at most, 95MW of access through access settlements.

However, regional settlement is similarly loss-adjusted: a generator is paid the product of its output, the RRP and its MLF:

$$\text{Regional Pay\$} = G \times (MLF \times RRP) = (G \times MLF) \times RRP$$

So, if the above generator is dispatched at 100MW it only gets paid, at RRP, for 95MW.

In this sense, loss-adjusting access is consistent with the "compensation" model of OFA. If the generator above is constrained off it really only loses 95MW of access to the RRP and, by purchasing 100MW of access, it is fully compensated for that loss. If there were no loss-adjustment, and the generator were provided with the full 100MW of access, it would be *over*-compensated.

Access Certainty

MLFs are re-set annually. The generator in the example above may find its MLF is reduced to 0.90 at the reset. In this case, its access level is reduced to 90MW and so is its regional settlement payment.

Thus, loss-adjusting access creates some uncertainty for the generator. But this uncertainty is mirrored in regional settlement: any changes to MLF will similarly impact the amount of generation sold at the RRP. Each generator will already have processes for managing the latter uncertainty: for example, adjusting its hedging position to track the annual MLF variation. It can take similar actions in relation to its access.[89]

Therefore, although loss-adjustment creates some access uncertainty for generators, this is likely to be modest compared to the corresponding RRP uncertainty and can be managed using similar mechanisms.

FAS

As discussed in the next chapter, FAS sets a target for flowgate capacity equal to the aggregate of target firm entitlements on a flowgate. Since target access (and so target entitlements) is loss-adjusted, TFGX will vary annually in response to MLF variation.

However, since flowgate capacity is defined as total usage (on congested flowgates), and usages too are loss-adjusted, flowgate capacity will similarly vary in response to MLF variation. The variations may not be exactly in sync: TFGX will vary based on the MLF variation of *firm* generators whilst FGX will vary based on the MLF variation of *dispatched* generators. However, it seems likely that loss-adjusting entitlements will reduce, rather than exacerbate, the impact of MLF variations on FAS obligations.

Conclusion

The discussion above identifies some conceptual benefits form loss-adjusting access. It also seems intuitive to treat usage and entitlements similarly, since access settlement payments are based on the difference between them.

4.3.10 Measuring Generation

Overview

As explained in section 4.2.5 it is proposed to measure generation output for in access settlement using a sent-out (SO) approach. Alternative approaches are to use as-generated (AG) or metered quantities.

There are a number of factors to consider in making this design choice:

- dispatch constraint formulation;
- pricing of auxiliary load;
- measurement practicalities;
- transmission planning; and
- access pricing.

---

89 For example, buying short-term firm access as described in section 7.2.4.

These factors are considered in turn below.

Dispatch Constraint Formulations

The generator variables in the dispatch constraints used by NEMDE represent AG output. If access settlement is based on a different measure, there may be a need to adjust the variables extracted from the dispatch constraints accordingly.

Only participation factors are extracted from constraint equations. (As explained previously, FGX is then calculated based on total usage, rather than being based on the constraint RHS.) There may be some modest difference between AG-related participation factors and SO-related participation factors[90] but this is not expected to impact materially on access settlement.[91]

Pricing of Auxiliary Load

Recall that SO output incorporates auxiliary load:

SO output = AG output – auxiliary load

With access settlement based on SO output, auxiliary load is, in effect, paid for at the local price rather than at RRP. Since the local price will be lower than RRP during congestion, there is an incentive for a generator to maximise the scope of load that is treated as auxiliary for the purposes of access settlement.

It is proposed that generator discretion in this area is restricted through the application of clear principles and guidelines about which load can be classed as auxiliary.

Depending upon meter location and configuration, metered output[92] can also include some auxiliary load. Using metered output for access settlement could encourage generators to reposition their meters in order to maximise the implied auxiliary load. Moving meters is expensive, and incurring such costs would be inefficient, since there would be no benefit to the market as a whole. It is for this reason, that metered output is not a suitable basis for access settlement.

Measurement Practicalities

Typically, AG generation is not metered by revenue quality meters, only by SCADA meters. The specification and supervision frameworks for SCADA metering are not intended to make it sufficiently accurate for financial settlement. Thus, if access

90 Because auxiliary load can be correlated with generation output.

91 Because generators are only paid RRP on their SO output, they are likely to adjust their offer prices to account for any difference between AG-based and SO-based participation. This adjustment will broadly counteract the impact on access settlement of using AG-based participation factors in tandem with SO-based output quantities.

92 Ie, the quantity measured by the existing unit meters that are used for settlement, as described in appendix B.1.

settlement were based on AG output either new meters would need to be installed or the SCADA metering processes would need to be enhanced. This is likely to be expensive. In comparison, SO output can be reasonably estimated based on existing revenue metering.

Transmission Planning

As discussed in section 4.2.1, flowgate capacity depends upon local demand. A TNSP that is required to provide TFGX under FAS must be able to estimate the level of local demand in order to decide how much network capacity to provide.

If access settlement is based on AG output, then all auxiliary load effectively becomes local demand. A TNSP must then estimate auxiliary load in order to plan network expansion. For example, suppose a 100MW (in terms of AG output) generator procures 100MW of firm access. If that generator had 5MW of auxiliary load, the TNSP would have to provide an extra 95MW of network capacity. If, on the other hand, the generator had 40MW of auxiliary load (eg, because it is being developed as a co-generator), the TNSP only has to provide an extra 60MW of network capacity.

Therefore, using AG output increases planning uncertainty for TNSPs, compared to using SO output.

Access Pricing

The price for 100MW of firm access, in the example above, will be based on the access pricing model. This model does not incorporate auxiliary load and the generator would be charged the same irrespective of its auxiliary load. In the context of using AG output, this would be unfair, given that the cost impact on the TNSP very much depends upon auxiliary load.

If using SO output, the generator with the higher auxiliary load would need to purchase less firm access and would be charged a lower price. This is a more efficient outcome. On the other hand, this re-emphasises the importance of restricting generator discretion on what is classed as auxiliary load.

Conclusion

To conclude, using SO output for access settlement is likely to be simpler and cheaper. It aligns better with the OFA principle that generators are charged – through access settlement or access pricing – based not on how much they *produce* electricity but on how much they *use* the shared transmission network.

On the other hand, an SO approach encourages generator to maximise the scope of their auxiliary load and it is important that generator discretion in this area is appropriately restricted.

5 Firm Access Standard

5.1 Overview

The *quality* (ie, the firmness) of the firm access service is predicated on the capacity and reliability of the shared transmission network that underpins it. Thus, two factors are required to provide generators and FIR holders with confidence that service quality will be maintained: a *service standard* that specifies the minimum service quality that must be provided to *each* firm participant; and a corresponding *network standard* that specifies the level of network capacity that the TNSP must build and deliver to provide the minimum service quality to *all* firm participants. The *Firm Access Standard* (FAS) performs both of these roles.

The FAS consists of a planning component and an operational component.

The Firm Access *Planning* Standard (FAPS) requires that a TNSP plans to provide sufficient network capacity such that each firm participant receives its registered access under specified *FAPS conditions*: the generation, demand and transmission conditions under which firm access is likely to be of highest value to the holder.

The Firm Access *Operating* Standard (FAOS) requires that a TNSP makes this planned network capacity available *operationally* in an efficient way: by trading off the cost of operation against the cost, to firm participants, of any shortfalls in network capacity. The FAOS is achieved by establishing an operational incentive scheme, under which efficient TNSPs are rewarded and inefficient ones penalised. The scheme will be established and overseen by the AER.

5.2 Design Blueprint

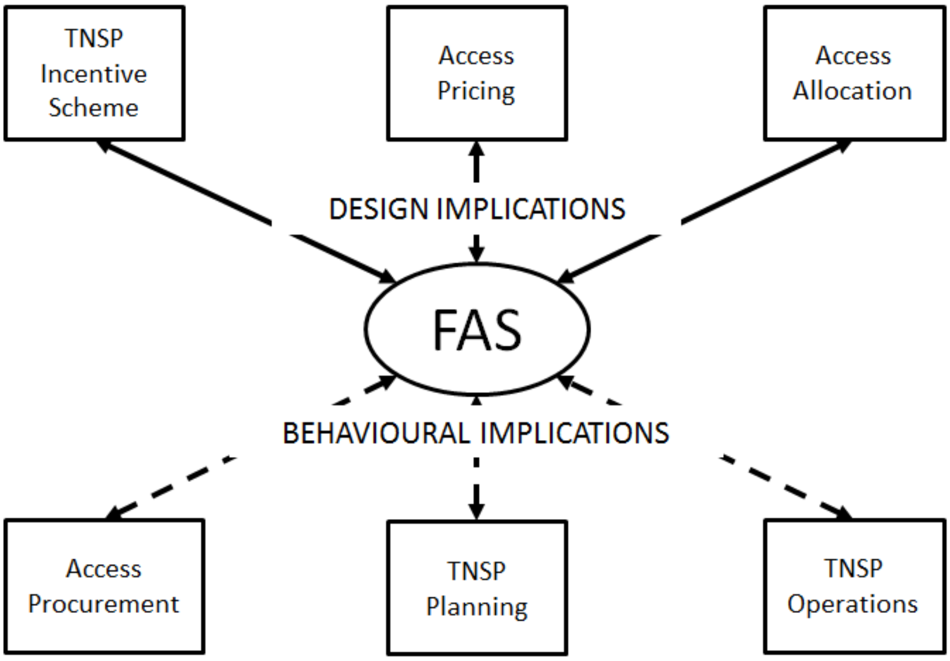
5.2.1 Role of the FAS

Overview

The firm access standard (FAS) specifies the firmness of the firm access service and places obligations on TNSPs to provide sufficient network capacity to deliver that firmness. The target network capacity set by FAS reflects the aggregate amount of firm access that is registered. Each new firm access registration, therefore, increases the amount of network capacity that a TNSP must provide. In this way, the FAS is the vehicle through which market demand for firm access drives TNSP planning and operations.

These linkages are illustrated in Figure 5.1 below.

Figure 5.1 Role and context of FAS



Target Flowgate Capacity and Shortfalls

The concepts of target flowgate capacity (TFGX) and effective flowgate capacity (EFGX) were introduced in chapter 4. TFGX is critical to determining the level of access received:

1. If EFGX equals or exceeds TFGX, on all flowgates, then all generators and DICs will receive effective access at least equal to their target firm access.[93]

2. If EFGX is less than TFGX, on at least one congested flowgate, then effective access will be below the target firm access level for at least one firm participant.

When EFGX is *less* than TFGX on a flowgate, there is said to be a (flowgate) *shortfall* equal to the difference. The FAS specifies obligations on TNSPs in relation to these shortfalls. To avoid or remedy shortfalls, a TNSP may have to take actions to increase EFGX: for example, build new transmission capacity.

TNSP Planning and Operations

The OFA model treats differently the processes of:

- *transmission planning*: the design and development of new transmission assets to maintain or expand network capacity; and

93 Recall that the target firm access is the lower of registered access and generator capacity.

- *transmission operation:* the operation of these assets (eg, maintenance scheduling), and other ancillary processes, to ensure that network capacity is available when required for generation dispatch.

Planning is characterised by long timescales (several years) and expensive, "lumpy", long-lived assets: eg, transmission lines and transformers. Operation is primarily about making the best use of the assets that have been developed in the planning process. Any capital expenditure involved in operations has low costs and fast payback times compared to planning. Although the planning-operations dichotomy is somewhat artificial – there are grey areas which do not clearly fall into either category – it is already used in the NEM in relation to TNSP revenue regulation and is suitable for use in the OFA model.

In accordance with this dichotomy, the FAS consists of two components:

- a planning component: referred to as the *Firm Access Planning Standard* (FAPS); and

- an operational component: referred to as the *Firm Access Operating Standard* (FAOS).

Access Pricing

The FAS expresses the amount of flowgate capacity that must be provided in terms of the amount of registered access. Thus any *new* firm access imposes *additional* costs (either immediately or in the future) on a TNSP by increasing the amount of flowgate capacity that it must provide. This cost is reflected in the access price that is charged for firm access. Thus, the pricing method must reflect the FAS. This is discussed further in section 6.2.1.

5.2.2 FAPS

FAPS Definition

The FAPS requires that a TNSP undertakes transmission *planning* so that, under specified *FAPS conditions*, there are no shortfalls to be expected on those flowgates that could become *congested* operationally.

There are several aspects to this definition:

- the definition of the FAPS conditions;

- the meaning and calculation of effective flowgate capacity in a *planning* context;

- the identification of flowgates that could be congested;

- the meaning and enforcement of a *planning* obligation as opposed to an operational obligation: ie, the TNSP must *plan* to deliver the flowgate capacity as opposed to *actually* delivering it; and
- the allocation of responsibility between TNSPs where access is provided jointly: eg, in relation to inter-regional access.

These are discussed in turn below.

FAPS Conditions

The FAPS conditions will specify annual FAPS *snapshots* of market and transmission conditions:

1. the level and distribution of consumer demand;
2. the availability and ratings of network assets;
3. the output of scheduled, semi-scheduled and non-scheduled generation; and
4. any other factors which materially affect flowgate capacity and congestion.

The FAPS conditions will reflect those conditions under which *market participants value access most highly*. This will generally be the case under peak demand conditions since:

1. regional prices are likely to be high; and
2. obligations on retailers to supply customers, and on generators, to back hedging contracts with retailers, are likely to be high.

An example specification of FAPS conditions is described and discussed in appendix C.1. The actual specification would be developed as part of the implementation and operation of the OFA.

FAPS Capacity

The effective flowgate capacity of a flowgate under FAPS conditions is referred to as the *FAPS capacity*. Thus, the FAPS requires that FAPS capacity exceeds TFGX.

As discussed in section 4.2.2, the effective flowgate capacity is given by the formula:

EFGX = FGX + FGS

Where:

EFGX = effective flowgate capacity

FGX = flowgate capacity

FGS = flowgate support: when flowgate support generators are dispatched

The EFGX of a thermal flowgate will therefore depend upon the following factors:

- the thermal rating of the transmission element that the constraint protects;
- the topology of the transmission network;
- the level and distribution of demand;
- the location and output of any non-scheduled generation; and
- the output and participation of (scheduled) flowgate support generators.

Assumptions for these variables would be determined based on the FAPS specification. For example, if it specified summer peak conditions, then the ratings would reflect peak summer temperatures, the demand would be based on peak forecast demand, and so on.

The assumed commitment and output of scheduled and semi-scheduled generators, including flowgate support generators, would similarly be predicated on the FAPS specification. For example, if the FAPS specified a high-RRP condition, it could be assumed that all scheduled and semi-scheduled generators would aim to run at their full availability,[94] although the implied aggregate generation may then have to be reconciled against the aggregate demand.

Generally, it would *not* be appropriate to assume that generators all operate at their registered access level since this would exclude, for example, the flowgate support that might be provided by non-firm generators. This issue is discussed further in section 5.3.2.

The EFGX for stability constraints, in addition to the factors listed above, may depend upon:

- the commitment status of scheduled and semi-scheduled generators: ie, whether they are operating and connected to the shared network; and
- the availability and settings of TNSP-operated power system control equipment: such as capacitor banks.

Commitment of scheduled and semi-scheduled generators would be based on their assumed output under FAPS conditions, discussed above.

It would be reasonable to assume that control equipment is set so as to maximise flowgate capacity. However, some conflicts may arise in doing this: eg, the setting that optimises capacity of one flowgate may lead to sub-optimal capacity on another flowgate.

In summary, the FAPS conditions need to specify enough detail so that the assumed status of transmission, generation and demand can be forecast.

---

94 Unconstrained Intermittent Forecast Generation (UIGF) for semi-scheduled generators.

Flowgates That Cannot Become Congested

As discussed in section 4.2.1 every transmission element (line or transformer) is a flowgate. Should the FAPS require that a TNSP provides TFGX on every one of these flowgates? Firm generators will only be affected by a flowgate shortfall when the flowgate is congested. But is it possible to identify, in planning timescales, flowgates that cannot *possibly* be congested? A flowgate becomes congested when total usage reaches flowgate capacity. So these two aspects need to be considered.

For an individual generator, the participation factor represents the usage per unit of output. Therefore, congestion of a flowgate is most likely when high participation generators are dispatched. Operationally, in each region, total generation plus total interconnector imports must equal regional demand. Therefore, the *maximum usage* on a flowgate arises when generators and DICs are dispatched in order of descending participation factor (ie, highest participation first) up to the regional demand level.

Flowgate capacity depends upon the availability and ratings of network assets, as well as the level and location of local demand and non-scheduled generation. *Minimum flowgate capacity*, for a particular demand level, would then be the flowgate capacity in the most adverse conditions.

The *headroom* on a flowgate is then the lowest value of the difference between minimum flowgate capacity and maximum usage, calculated under various demand levels. If the headroom is greater than zero, then the flowgate cannot become congested and so FAPS can be deemed to be met for that flowgate, even if EFGX is less than TFGX.

Amongst other things, the headroom will depend upon what network conditions are considered. Clearly, flowgate capacity will be lower when there are network outages. Including these conditions in the headroom assessment might then prompt expansion of a flowgate that could, at worst, become congested only rarely. On the other hand, if outage conditions are not considered, and so flowgate expansion is not required, firmness will be reduced.

In conclusion, the FAPS will need to specify the conditions – especially network conditions – that are assessed in the headroom calculation.

Planning Processes and Transparency

The FAPS is only an obligation to *plan*. The occurrence of a shortfall *operationally* is not necessarily a breach of FAPS, so long as it is not attributable to a planning failure: ie, there were no actions that the TNSP could and should have taken in planning timescales that would have prevented the shortfall occurring.

There are several possible causes of operational shortfalls occurring:

1. the operational conditions are different to FAPS conditions: this is generally the case of course, since the FAPS conditions will be highly specific and likely to occur at most once per year;

2. the relevant flowgate was not identified during planning: for example, an unexpected stability constraint has arisen;

3. the relevant flowgate was identified but was expected not to become congested;

4. the flowgate was identified but the EFGX was over-estimated in the planning process: eg, due to demand forecasting errors;

5. the flowgate was identified but the TFGX was under-estimated: eg, because new firm access was issued after the planning had taken place; and

6. the flowgate shortfall was correctly identified but the necessary flowgate expansion did not occur: eg, due to delays in the completion of an expansion project.

The first cause does not imply a problem with FAPS compliance, unless the operational conditions are very similar to FAPS conditions. The fifth cause should never arise in practice, due to restrictions placed on the issuance process, as described in chapter 7. Shortfalls caused by remaining factors should be able to be minimised through an effective planning process in accordance with FAPS.

Therefore, the TNSP planning process should:

- establish accurate forecasts of the factors that drive FAPS capacity: relating to demand, generation and transmission;
- comprehensively identify and analyse all potentially-congested flowgates;
- identify shortfalls on potentially-congested flowgates; and
- identify, select and develop flowgate expansion projects to remedy these shortfalls.

To facilitate monitoring, each of these sub-processes should be transparent. This would be through the TNSP providing the necessary information to the market and through the AER: primarily through inclusion in the Annual Planning Reports that TNSPs will continue to publish under OFA.

Flowgate Tagging

The NEM transmission network is shared and interconnected across regions. It may not always be clear in which region a flowgate is located. Even when it is clear, it is not necessarily the case that the local TNSP should be responsible for maintaining FAPS with respect to that flowgate.

For example, a transmission constraint in dispatch typically prevents thermal overload or power system instability following a contingency. The overload/instability and the contingency may be located in different regions. So which is the relevant region for assigning FAPS responsibility?

Assigning responsibility for a flowgate is informally referred to as *flowgate tagging*: each flowgate is tagged with the region of the responsible TNSP. Possibly, in some cases, responsibility might be shared and this would be reflected in the flowgate *tag*.

The FAS will need to specify how flowgate tagging is undertaken. The tags apply to both FAPS and FAOS.[95] Notwithstanding that a flowgate may be tagged to a single TNSP, its EFGX may depend upon the actions or assumptions of another TNSP. Thus, planning will need to be coordinated across regional boundaries: this requirement reflects a continuation of current practice.

FAS responsibility is based on planning responsibility rather than ownership or operation. There are some TNSPs who may own or operate assets but who do not have planning obligations in that region. These would *not* have FAS obligations or responsibilities; the name of the planning TNSP in the region would, instead, be tagged to the related flowgates. The planning TNSP, however, is likely to make arrangements with the asset owner or operator, to ensure that FAS is able to be maintained.

5.2.3 FAOS

Overview

Under the FAOS, each TNSP will be subject to an *operational incentive scheme*. The incentive scheme will have the following elements:

1. Incentives will be based on *capped shortfall costs*.
2. An *annual benchmark* for aggregate capped shortfall costs will be set by the AER: TNSPs will be rewarded or penalised based on the difference between benchmark and actual capped shortfall costs.
3. TNSP rewards (or penalties) will be paid (or received) by firm generators and DICs. Settlement will occur in the following year.

The *shortfall cost* is the product of the shortfall (in MW) and the flowgate price. When there is a shortfall, entitlements to firm generators are reduced, in aggregate, by the amount of the shortfall. This means that access payments to firm generators, in aggregate, are reduced by the amount of the shortfall cost.[96]

95 Discussed in section 5.2.3.

96 Recall that access settlement payments are based on the formula: (E-U) x FGP. So every Q MW reduction in entitlement causes Q x FGP reduction in access settlement payments.

FAOS Objective

Unlike the FAPS, the FAOS does not specify a particular level of flowgate capacity that a TNSP must provide. Rather, the FAOS has an objective of *efficient* transmission operation.

TNSPs incur the *direct* costs of transmission operation. Firm generators and DICs bear the *indirect* costs – the shortfall costs - that arise when operational decisions lead to flowgate shortfalls, causing firm access to be scaled back and access settlement payments to be reduced. Transmission operation is *efficient* when the sum of the direct and indirect costs is minimised.

For example, shortfalls may occur during scheduled outages of transmission assets. The shortfall cost can be reduced by scheduling the outages for periods when congestion is low and also by minimising the outage duration. However, this may entail some extra cost to the TNSP. The FAOS objective is for there to be an efficient trade-off between these two costs.

The FAOS could, then, be stated simply as: "the TNSP must be efficient in transmission operation". But this would be practically unenforceable since it is impossible to know exactly what is efficient and what is not. Instead, the FAOS incentive scheme creates *incentives* on the TNSP to operate efficiently and TNSPs are expected, rationally, to respond to these incentives.

Shortfall Costs

The incentive scheme provides incentives for efficient transmission operation by charging a TNSP for the capped shortfall costs that arise during periods of flowgate shortfalls.

In principle, if a TNSP is charged the full amount of shortfall costs, then it will take these into account in its operational decisions. It would face the shortfall costs and its own internal costs *and* would rationally seek to minimise the sum of the two. It would, therefore, endeavour to operate efficiently, as this is defined above.

But is it practically feasible to charge the *full* shortfall costs to the TNSP? The shortfall cost is the product of the shortfall amount and the flowgate price. Despite the TNSP complying with the FAPS, shortfall amounts could occasionally be in the hundreds of MWs at peak times: eg, due to forced outages of critical transmission elements during a demand peak.

Flowgate prices are related to the difference between the RRP and the offer price of the constrained-off generator. RRP can exceed $10,000/MWh at peak times, whilst offer prices are typically below $100/MWh and can be as low as -$1000/MWh.[97] Thus, flowgate prices can be in excess of $10,000/MWh at times.[98]

97 The market price floor.

98 They can be even higher, due to the gearing effect described in section 4.3.8.

Shortfall costs would, under such conditions, accrue at rates of millions of dollars per hour. Charging these shortfall costs in full would place extreme risks on TNSPs, far beyond what they face in the current regulatory environment, and so it is not practically feasible to charge the full shortfall costs to TNSPs.

Nested Caps

To mitigate this risk for TNSPs, it is proposed that the incentive scheme specifies caps on shortfall costs and only charge TNSPs up to the capped amount. Several, *nested* caps would be specified, applying at different timescales, for reasons discussed below.

It is common, in regulated incentive schemes, for there to be an annual cap: set as a percentage of annual revenue, for example. But if extreme shortfall costs were to arise, as described above, this cap could be hit very quickly, meaning no further penalties – and so no incentives – on TNSPs for the remainder of the year.

Alternatively, a settlement period (half-hourly) cap could be set. But if this is set too low, incentives on TNSPs to avoid high shortfall costs are substantially diluted. On the other hand, if it is set too high, there remains a possibility of a high, cumulative annual charge on the TNSP, meaning the TNSP continues to face a high risk.

A structure of nested caps provides some flexibility to help resolve these difficulties. An annual cap is set, which limits overall TNSP risk. This annual cap can be thought of as a *risk budget*. A series of caps is then set, at various shorter timescales. For example, two timescales might be used in conjunction with the annual cap as follows:

- a half-hourly cap of $50,000;
- a daily-cap of $250,000; and
- an annual cap of $15m.

Half-hourly shortfall costs are capped at $50,000. Cumulative daily capped shortfall costs are then further capped at $250,000. For example, suppose that, within a day, shortfall costs were $1m per half-hour for 10 hours. They would be capped initially at $50,000/half-hour. After 10 hours, these capped costs would have reached $1m. These costs would be capped at the daily cap of $250,000.

In the worst case, the daily-capped amounts could accumulate, day-after-day, to reach the annual cap after just 60 days. But this is unlikely: eg, if the extreme shortfall costs are caused by a forced outage, this is unlikely to last for so long. In any case, once the annual cap is reached, no further costs are charged to the TNSP. At the end of the year, if the annual cap is exceeded, the capped shortfall costs in each settlement period are all scaled back by the amount needed to bring the cumulative cost back to the annual cap.

Nested caps are discussed in more detail in appendix C.2.

Annual Benchmark

The incentive scheme should not penalise a TNSP that operates efficiently. The TNSP should be able to recover its efficient operating costs: both its direct costs and the capped shortfall costs. An annual benchmark is specified in the incentive scheme which provides for this cost recovery.

The benchmark is set by the AER, based on its assessment of these efficient costs. Some operating costs will already be allowed for in the AARR[99] and so would be excluded from the benchmark.

For example, suppose that, over a year:

- efficient shortfall costs are $10m;
- efficient capped shortfall costs are $6m; and
- additional TNSP costs from efficient operation are $3m.

Then the total costs borne by an efficient TNSP are $9m ($6m plus $3m) and so the annual benchmark is set at this level. The TNSP is penalised the capped shortfall costs. So, if the efficient outcome occurs, the TNSP is paid $9m (its annual target), pays back $6m (the capped shortfall cost) and is left with $3m, which covers its operating costs.

The TNSP may, *inefficiently*, decide to spend no money on operations (over and above that allowed for in the AARR) and, as a result, capped shortfall costs blow out to $15m, say. The annual benchmark remains at $9m, so the TNSP is $6m worse off in this case.

Alternatively, the TNSP might inefficiently *over*-spend on operations: eg, spending $10m and bring capped shortfall costs down to $4m. With costs of $14m ($10m plus $4m) this overspending TNSP is $5m worse off than the efficient TNSP.

It is not known, in practice, what the level of these efficient costs might be, so the AER is likely to adopt an approach that allows the efficient cost level to be progressively revealed over time. For example, the annual benchmark might be partly based on historical outcomes. Although these outcomes may reflect inefficient operation, and thus lead to a generous annual benchmark, the incentives on the TNSP to improve efficiency remain. If the TNSP can *improve* its operating efficiency compared to the historical period, it will make a windfall gain. Over time, as this more efficient operation is reflected in the historical outcomes, the annual benchmark will become more challenging and the windfall gains will reduce. The benchmark should converge to the efficient level.

Annual Settlement

Since the benchmark is set on an annual basis, and the shortfall costs have an annual cap, the amount payable or receivable by the TNSP under the incentive scheme is not

99 Described in section 8.2.2.

known until the end of the year. For this reason, settlement of the incentive scheme does not occur until the following year.[100]

The gains from efficient TNSP operation flow to firm participants (ie, firm generators and firm DICs), who receive a firmer level of access as a result. For this reason, the incentive scheme is settled with these parties:

- if capped shortfall costs for the year exceed the annual benchmark, the TNSP pays the difference to the firm participants; and
- if capped shortfall costs for the year are less than the annual target, the difference is paid by firm participants to the TNSP.

The aggregate of payments from (or to) firm participants in a region must equal the payment to (or from) that region's TNSP. The payment to or from each DIC would be allocated between FIR holders on that DIC, in proportion to their registered access.

To reduce cashflow impacts on TNSPs and firm participants (with whom settlement occurs) it may be preferable to spread settlement across the following year rather than in a single lump sum. For example, one twelfth of the annual settlement amounts could be payable in each month of the following year.

Payment to Firm Participants

A method is needed to determine an appropriate allocation of TNSP payments between firm participants. The objective is to ensure that no individual participant is worse off as a result of the incentive scheme: in particular, any payment that a participant makes into the scheme should be less than the benefit obtained - from lower shortfall costs – as a result of the scheme being in place.

Suppose that, in a year, the capped shortfall costs are below the benchmark, so that the TNSP is rewarded. Although, the benefits of lower than benchmark annual shortfall costs would accrue to the firm participants *collectively*, the gains may not be shared equally. It is possible that some firm participants have, notwithstanding the collective position, suffered a *worse* than benchmark year. It would be unreasonable to expect these participants to also contribute to the TNSP reward.

Conversely, in a year of higher than usual benchmark shortfall costs, some participants may have nevertheless had a good year and do not need to be compensated by receiving a share of the TNSP penalty payments.

Therefore, the allocation should, ideally, be based on a comparison, for *each* firm participant, of actual capped shortfall costs against benchmark shortfall costs. For example, the payment to each firm participant could be set *equal* to the difference between their actual and benchmark capped shortfall costs. The total payment is then the difference between the total actual and total benchmark capped shortfall costs

100 It would be possible to have settlement during the year, with continuous reconciliation as further shortfall costs are incurred, but this increases complexity for no obvious benefit.

which equals the amount payable by the TNSP. So, incentive scheme settlement balances under this approach.

An example of this is presented in Table 5.1 below.

Table 5.1 Incentive scheme settlement example

$m	TNSP	Firm Participant 1	Firm Participant 2	Firm Participant 3
Benchmark costs	15	5	4	6
Actual Capped Shortfall Costs	10	2	5	3
Settlement	5	3	-1	3

Attributing Shortfall Costs between firm participants

If settlement is to be based on shortfall costs incurred by individual firm participants, as described above, these amounts need to be defined and calculated.

It is straightforward to attribute *uncapped* shortfall costs to each participant. These costs are defined, for each flowgate and each settlement period in which a shortfall occurs, by the formula:

> Attributed uncapped shortfall cost = FGP × (target firm entitlement – actual firm entitlement)

Capping the shortfall cost is equivalent to *scaling* the shortfall cost, so long as the scaling factor is set correctly. For example, if shortfall cost in a settlement period is $100,000 but the cap is $20,000 then a scaling factor of 0.2 is equivalent to the capping. This scaling factor can be applied to attributed shortfall costs:

> Attributed capped shortfall cost = scaling factor x attributed uncapped shortfall cost

With nested caps there are multiple scaling factors applied, but the same principle can be used for allocating capped shortfall costs. The detailed algebra for this is presented in appendix C.3.

If the annual benchmark capped shortfall cost is based on historical or simulated outcomes, a similar approach can be taken to attributing the benchmark cost between firm participants. These benchmark costs could not be used *directly*, since some firm access may have expired or been procured since the historical period. However, they can be used to develop some simple formula for allocating the historical benchmark costs to *current* firm participants, based on the level and location of firm access.

The benchmark operating costs could be allocated between firm participants based on a simple measure: eg, in proportion to registered access.

Cross-regional Issues

As discussed in section 5.2.2, flowgates will be tagged with the region of the responsible TNSP. These same tags[101] are used in the FAOS incentive scheme. The shortfall cost accruing on a particular flowgate will be attributed to the TNSP(s) to which the flowgate is tagged. Tagging will also be factored into the target-setting process, based on the tags in the historical scenario.

A firm generator or DIC may participate in flowgates tagged to a TNSP other than their local TNSP. They would then need, in principle, to participate in incentive settlement with *multiple* TNSPs. This increases complexity and it needs to be considered whether this is worthwhile. For example, if 95% of a shortfall costs accrue to local[102] generators and DICs, cross-regional settlement may be unnecessary. In this case, the generator allocations of target and actual shortfall costs would need to be scaled up so that they still sum to the TNSP values.

5.3 Design Issues and Options

5.3.1 Is FAPS an Economic or Deterministic Standard

Introduction

Transmission reliability standards can be characterised as either *economic* or *deterministic*. An *economic* standard endeavours to explicitly value reliability and then requires that a TNSP *efficiently* maintains and expands its network in accordance with that valuation. The TNSP must follow the expansion plan that maximises net benefit: the benefit of improved reliability minus the expansion cost.

A *deterministic* standard, on the other hand, explicitly specifies the level of redundancy that a TNSP must provide, such that load can continue to be supplied even when the specified number of transmission elements are out of service. The TNSP is then required to maintain that standard at least cost.

There are pros and cons associated with each type of standard and it is beyond the scope of OFA to consider which is preferable. But it is useful to consider how the FAPS might be categorised. This affects the way that FAPS is perceived by stakeholders. It also has some implications for OFA design which are discussed in later chapters.

It is argued in this section that FAPS, in a sense, falls into both categories.

---

101 Or, at least, the same tagging *principles*. The tagging *process* will be different. Under FAPS, the FAPS constraints are tagged, by the TNSP. Under FAOS, the NEMDE transmission constraints are tagged, probably by AEMO.

102 Ie, based in the same region as the flowgate.

Deterministic Definition

FAPS is *expressed* in a very similar way to a deterministic standard. TNSP must deliver a *specified* amount of network capacity under *specified* conditions. The value of the capacity – and the access it provides – is not considered in the standard and so there is no opportunity for a TNSP to decide *not* to provide the specified capacity, on the basis that the cost of the associated expansion exceeds the value of the access provided.

This is in contrast to the FAOS, which is not expressed deterministically and, under which, a TNSP is expected to make efficient trade-offs between cost and value.[103]

Economic Delivery

But whilst the FAPS is *defined* deterministically, it is nevertheless designed to *deliver* an economic standard. This is because the level of capacity that the TNSP must provide is predicated on the amount of firm access issued which, in turn, is predicated on how generators value firm access. So, the value of firm access is, indirectly, incorporated into expansion planning.

A generator will procure additional firm access where its value (as perceived by the generator) exceeds the access price. In turn, the access price reflects expansion costs. So, expansion will occur only when a generator procures firm access, which only occurs when the access value exceeds the associated expansion cost. In short, expansions should only occur when they deliver net benefit, similar to an economic standard.

The FAPS is superior to a conventional economic standard in one respect. Expansions are predicated on generator valuations rather than TNSP valuations. Since the generator is paying the bill, it is best placed to decide whether firm access provides value for money. In contrast, under an economic reliability standard, it is the TNSP who estimates the value of reliability, rather than the consumer (who is ultimately paying the bill).[104]

This feature of the FAPS is predicated on the *optionality* of firm access. Generators have the right to "walk away", opting for non-firm access instead, if they consider that the access price too high. A *mandatory* firm access regime, in contrast, would deliver a *deterministic* standard of access.

Conclusion

Despite being expressed deterministically, the FAPS delivers an economic standard of access. In this respect, it is consistent with FAOS.

103 As discussed in section 5.2.3.

104 Although TNSPs typically base their estimates of reliability value on customer surveys.

5.3.2 A Dispatch Analogy for FAPS

Firm Generation Dispatch

Define a *firm generation dispatch* as meaning that *every generator is dispatched at its target firm access level.*[105] Suppose, for simplicity, that there are no flowgate support generators. Then, under this dispatch, the usage of a generator on a flowgate is:

$$U = \alpha \times \text{target firm access} = \text{target firm entitlement}$$

Aggregating across all generators participating in the flowgate:

$$\text{Total } U = \text{aggregate target firm entitlement} = \text{TFGX}$$

If this dispatch is *feasible* then all flowgate constraints are complied with: ie, on every flowgate:

$$\text{Total } U \leq \text{FGX} \tag{5.1}$$

Putting all of these equations together we have, on every flowgate:

$$\text{FGX} \geq \text{TFGX}$$

Because there is no flowgate support, EFGX is equal to FGX. So this inequality is identical to the FAPS requirement. Thus: *in the absence of any flowgate support generation, the FAPS requirement is equivalent to requiring that a firm generator dispatch is feasible under FAPS conditions.*

Flowgate Support

How does this principle change when there are flowgate support generators? In this case, total usage includes the usage of flowgate support generators, and so:

$$\text{Total } U = \text{TFGX} - \text{firm flowgate support}$$

Where firm flowgate support is the flowgate support provided by firm generators being dispatched. Adjusting equation 5.1 accordingly gives:

$$\text{TFGX} - \text{firm flowgate support} \leq \text{FGX}$$

So:

$$\text{TFGX} \leq \text{FGX} + \text{firm flowgate support}$$

However, EFGX under FAPS conditions depends upon the actual flowgate support provided under FAPS conditions:

---

105 Recall that target firm access is the lower of generator capacity and registered access.

$$EFGX = FGX + \text{actual flowgate support} \quad (5.2)$$

Firm flowgate support will differ from the actual FAPS flowgate support because:

- it does not include any flowgate support from non-firm generators; and
- it may overstate the flowgate support from firm generators: ie, if they were not be expected to be fully available and dispatched under FAPS conditions.

Therefore, the inequality above is not the same as the FAPS requirement. By implication, FAPS is *not* equivalent to the feasibility of firm generation dispatch under FAPS conditions. To assess FAPS compliance correctly, flowgate support needs to be calculated explicitly, based on the FAPS snapshot.

Nevertheless, thinking of FAPS in terms of a feasible firm generation dispatch is a useful and intuitive analogy. This analogy plays a key role in access pricing.

Demand Balancing

Strictly speaking, the firm generation dispatch will always be infeasible; aggregate firm generation is unlikely to exactly match aggregate FAPS demand, so demand and generation will be out of balance. Conceptually, *notional demand* or *notional generation* can be added at the RRN to correct this imbalance.

The flowgate inequalities presented above are unaffected by this addition because:

- generation at the RRN has zero usage on all flowgates;[106] and
- demand at the RRN does not affect FGX.[107]

Fictitious Flowgate Congestion

In a situation where aggregate firm generation is much higher than aggregate FAPS demand, the large amount of notional demand added to the RRN makes the firm generation dispatch look rather implausible and unrepresentative. Congestion might occur, under the firm generation dispatch, on flowgates that could never be congested in reality: in particular, those flowgates close to the RRN which are supplying the high notional demand.

As discussed in section 5.2.2, FAPS does not require that TFGX is maintained on flowgates that *cannot* be congested operationally. Thus, in this respect, feasibility of the firm generation dispatch is a sufficient, but not a necessary, condition for FAPS

---

106 By definition, a flow from the RRN to the RRN does not use the network and so makes no use of thermal flowgates. It is implicitly assumed that this notional generation does not affect power system stability and so has no impact on stability flowgates.

107 Recall, however, from section 4.2.1 that local demand – located at nodes other than RRN – does change FGX. So, it is important that the notional demand is only added at the RRN.

compliance. Apparent congestion (in the firm generation dispatch) on some flowgates local to the RRN can potentially be ignored for planning purposes.

Conclusion

The feasibility of firm generation dispatch under FAPS condition is a reasonable approximation to the FAPS requirement. It provides intuitive understanding of what the FAPS requirement means and is also the basis for access pricing.

Nevertheless, it is only an approximation and should not be relied upon for planning purposes. In particular:

- the flowgate support provided under FAPS conditions may differ from that implied by a firm generation dispatch; and
- the addition, in the firm generation dispatch, of notional demand at the RRN could create fictitious congestion on flowgates close to the RRN.

5.3.3 Super-Firm Access

Need to Reflect Super-firm Access in FAPS

The option of incorporating a super-firm access service into the OFA design was discussed in section 4.3.6. Through the provision of super-firm entitlements, super-firm generators would obtain a higher level of access than firm generators under shortfall conditions. Other things being equal, this implies that access firmness for firm generators would be diluted somewhat if the super-firm service were introduced.

This dilution effect can be corrected if FAPS is adapted to reflect super-firm access, such that a higher level of EFGX is required. The super-firm access is then backed by this additional network capacity, rather than being borrowed from firm generators.

Two changes are needed to the FAPS:

- a revised definition of TFGX; and
- a revised identification of uncongested flowgates.

These changes are discussed in turn below.

Adjustment to TFGX

Super-firm access must be incorporated into the defniition of TFGX, as follows:

TFGX = total target firm entitlements + total target super-firm entitlements

For example, if a 100MW super-firm generator, with 120MW of registered access, is the sole participant in a radial constraint, then the TFGX will be 120MW, rather than 100MW.

Minimum Headroom

The calculation of flowgate headroom was discussed in section 5.2.2. A headroom greater than zero meant that the flowgate would not become congested and so no flowgate expansion is required under FAPS, even if EFGX is less than TFGX.

If a super-firm access service were introduced, this FAPS requirement could be strengthened, to require that the minimum headroom was no lower than the aggregate target super-firm entitlement on the flowgate. Therefore, a super-firm purchase could prompt some expansion on a flowgate with low, but positive headroom.

Implications

It might appear inefficient to require that a TNSP deliver "surplus" spare flowgate capacity on a flowgate that was rarely congested. But, in procuring super-firm access, the generator is asking the TNSP to provide this redundancy, and paying for the cost of doing so.

A super-firm generator is not concerned with its access under FAPS conditions. Rather, it is concerned with the amount by which it is scaled back under shortfall conditions. It is willing to pay for the network redundancy in order to be scaled back less during shortfalls.

5.3.4 FAPS Governance

Single FAPS snapshot

A TNSP must build a network that delivers target flowgate capacity under FAPS conditions. It is anticipated that the FAPS conditions would specify a *single*, annual snapshot. If, alternatively, FAPS specified *two* annual snapshots, say summer peak and summer overnight, the TNSP would have to build a network that meets the more stringent FAPS. It can obviously not respond by build two different networks: a "peak network" and an "overnight network".

On some flowgates, the overnight condition might be more stringent. Local demand can add to flowgate capacity and so overnight, when local demand is lower, effective flowgate capacity may also be lower. Shortfalls may appear first in the overnight condition and prompt expansion.

A generator may not really value or seek overnight access firmness: overnight RRPs are generally low. But the FAPS requires that it is provided nonetheless. And the cost of provision will be included in the generator's access price.

How could a generator respond? It would prefer a cheaper service that provides access certainty at peak but not overnight. But there is no such service available. There can only be one FAPS and one service standard predicated on it. It is not feasible for each generator to have its own FAPS. It could opt to go non-firm, but this would not deliver the access firmness that it required at peak.

To conclude, FAPS should specify the single annual snapshot under which generators are likely to value access most highly: or, possibly, specify multiple snapshots if these are likely to be also valued sufficiently highly.

Dynamic FAPS

The particular conditions under which access value is highest may vary over time. For example, some regions have only recently become summer-peaking and plausibly they could return to winter-peaking: if distributed solar generation gains significant penetration, say. A FAPS which *specified* summer peak FAPS conditions, say, could become anachronistic. The FAPS would need to be reviewed from time to time.

On the other hand, registered access will often have a long term: possibly decades. OFA processes – pricing and procurement – are geared to this long term. If a generator purchases 20-year firm access, it may be perturbed if the FAPS changes after 5 years, particularly if this impacts significantly on its access firmness. On the other hand, the access will have been priced based on the expected costs to comply with the then-current FAPS requirement over 20 years. If the FAPS changes, these costs may change and the TNSP or its customers will bear this variance.

FAPS governance needs to strike a balance between stability and adaptation. FAPS must be flexible enough to respond to changing market conditions but stable enough to avoid problems of "regret" for generators and TNSPs.

FAPS Governance

To achieve this balance, a possible governance framework is suggested below:

1. High-level *FAPS principles* will be described in the Rules.

2. The AER will develop *FAPS guidelines* that apply these principles to current market conditions.

3. TNSPs will specify the *FAPS conditions*. The AER will review these and approve them if they comply with the FAPS guidelines.

The AER will review its FAPS guidelines from time to time and amend them as needed to reflect changing market conditions. However, in doing so, the AER will have regard to the need to ensure stability in the FAPS requirement.

5.3.5 Reliability Standards

Introduction

As discussed in section 2.3.5, the FAPS has the *potential* to supersede the existing reliability standards in the role of maintaining reliability on the generation-side transmission network. However, it is *not* proposed to make any immediate changes to the existing standards, because it is prudent to leave them in place as a safety-net, given that the performance of OFA is uncertain.

Given this, the implications of maintaining the reliability standards and so having two overlapping standards, one market-driven and one centrally regulated, are worth considering.

Reliability Access

It is possible that, given a relatively low level of firm access, the reliability standard is more stringent than the FAPS. For example, suppose that, in a region there is:

- 10 GW of generation capacity in total;
- 5 GW of this is firm; and
- peak demand is 7 GW.

Using the dispatch analogy, FAPS only requires that 5GW of generation can be dispatched at peak: ie, under FAPS conditions. Reliability standards require that at least 7GW of generation can be dispatched; at least 2GW of this will inevitably be non-firm.[108] So 2GW of non-firm access is guaranteed to be provided under FAPS conditions. It seems that the fortunate non-firm generators will be provided with the same firmness of access as the firm generators, but will not have to pay for it.

That appearance is not entirely true. First, the similarity only applies in relation to planning. Operationally, firm generators will have the protection of the FAOS but the non-firm generators will not. Second, although – in the example – 2GW of non-firm access is provided, this is allocated amongst 5GW of non-firm generation capacity. No *individual* non-firm generator could be confident that it would receive the access.

To distinguish it from firm access, this 2GW of less-firm access that is provided in the example is referred to as *reliability access*. It raises some issues around pricing and procurement, which are discussed in the relevant chapters.

108 Although the conditions under which this reliability is required may be different to FAPS conditions, so the two standards are not directly comparable. For illustrative simplicity, it is assumed that the conditions are the same.

Inefficiency of Having Duplicated Standards

There is an argument that having duplicated standard is inefficient. If the FAPS is *more* strict than the reliability standards then OFA will lead to a higher level of transmission investment. But since the current level of investment is sufficient to maintain reliability, this additional investment is unnecessary and just creates additional costs in the market which will eventually be passed onto the consumer. On the other hand, if FAPS is *less* strict than the reliability standards then FAPS – and by implication OFA – has no impact on transmission planning and so there is no point in implementing it.

There are a couple of rebuttals to this argument. First, transmission planning currently is predicated on reliability standards but also on generator *location*. If generation-side transmission reliability is inadequate then a TNSP must expand access for an *existing* generator: ie, build new network capacity between the existing generator's location and the demand. OFA introduces new locational signals – through access pricing – that may change the location of new generation and so change reliability-driven transmission expansion plans, even if FAPS, *per se*, has no impact.

Second, reliability standards are not set in stone. Possibly, if confidence is gained that the FAPS is able to maintain generation-side transmission reliability, changes could be made to the reliability standards accordingly. Their role could be confined to the maintenance of demand-side transmission reliability, for example.

Conclusion

It is prudent and pragmatic to maintain the existing reliability standards alongside OFA and FAPS, at least initially. But it is not necessarily elegant or logical. To draw an analogy, suppose that jurisdictional responsibility for maintaining generation capacity margins remained in place at NEM commencement and so two drivers of generation expansion operated side by side: one market-driven, one centrally-regulated. This would necessarily create some friction and tensions.

The OFA design aims to minimise the frictions created by having parallel planning standards. In particular, when it is relevant to design decisions, the FAPS will be considered to be the primary driver of generation-side transmission reliability, with the reliability standard consigned to the role of "safety net".

5.3.6 FAOS Design

Economic Standard

As discussed in section 5.3.1, FAPS is a deterministically-expressed standard which delivers an economic standard because it is driven by the cost-benefit assessment that firm participants make when purchasing firm access.

FAOS, on the other hand, is expressed as an incentive scheme but is similarly expected to deliver an economic standard.[109] In this case, it is the TNSP that is making the cost-benefit assessment when deciding whether to incur additional operational costs to save on shortfall costs.

Firmness

Despite these similarities between FAPS and FAOS there is one fundamental difference. Generators can have reasonable confidence that their firm access will be delivered under FAPS conditions. But FAOS does not place any quantitative obligation on TNSPs at all. It is theoretically possible that a TNSP never delivers the firm access amount and yet nevertheless is "FAOS compliant".

In practice, this is unlikely. Given that the firm access must be delivered under FAPS conditions, it would be expected to be largely delivered under "FAPS-like" conditions: eg, if FAPS conditions are specified as being at summer peak demand, access would be typically be largely delivered under high (but not peak) summer demand.[110]

The different factors contributing to flowgate capacity are likely to have differing, intrinsic levels of firmness. Transmission assets will have some efficient availability factor: availability could perhaps be increased beyond this level by more intensive maintenance, but the extra costs involved may be uneconomic. Local demand will typically have the usual diurnal and seasonal pattern and will cause capacity on affected flowgates to vary similarly. Non-scheduled generation may be correlated with RRP and/or intermittent resources (eg, wind) and, again, will lead to corresponding variations in flowgate capacity.

Given these fundamental factors driving variations in flowgate capacity, it is plausible that – despite it not being specified – the firmness of access under FAOS is fairly predictable. The major unpredictable factor might be the extent to which there is spare flowgate capacity: ie, the magnitude of the difference between EFGX and TFGX. Lumpy expansions or firm access expiries could create significant spare capacity which will naturally lead to fewer shortfalls and so firmer FAOS. When making procurement decisions, generators might conservatively assume zero spare flowgate capacity and then estimate FAOS firmness based on the historical variations in flowgate capacity, driven by the factors discussed above.

Nested Caps

The FAOS design specifies the use of nested caps to apply to shortfall costs. The number, duration and level of these caps would be specified by the AER when it applies or resets each individual TNSP incentive scheme. What objectives should the AER have in mind when setting the caps?

---

109 Although, the nested caps mean that it is not necessarily strictly an economic standard.

110 Assuming that transmission conditions were also similar to those during the FAPS condition.

The annual cap determines the maximum financial downside that the TNSP is exposed to, which is likely to be the key risk criterion for the TNSP from a corporate financial viewpoint. If the annual cap is set such that the overall financial risk of the TNSP increases, this may have implications for the appropriate level of regulatory WACC which, in turn, will impact TUOS prices. Therefore, in setting the annual cap, the AER will be concerned with TNSP risk and regulation overall, rather than just FAOS and shortfall costs, *per se*.

The annual cap sets the *risk budget* and the shorter caps then allocate and distribute this risk budget through the year. The AER is likely to have two objectives when setting these shorter caps:

- to ensure that incentives remain on the TNSP for as much of the year as possible; and
- to cause the strength of the incentive to be commensurate with the ability of the TNSP to manage the shortfall costs, under different conditions.

If the annual cap is hit mid-year, there is no incentive for the remainder of the year. If the monthly cap is hit mid-month, there is no incentive for the remainder of the monthly, and so on. In such cases, the TNSP will bear no further shortfall costs and so has no financial incentive to minimise them.[111] Rather, the incentive is just to minimise its own operational costs.

The impact of this mid-period loss of incentives is likely to be limited for the shorter-timescale caps. For example, by the time that a TNSP realises that the trading interval-cap has been hit, it is too late for it to change its operational approach anyway. If it is dealing with a forced outage, say, it will continue working to return this to service so as to avoid incurring shortfalls costs in subsequent trading intervals. That may be true of the daily cap, too. On the other hand, once the weekly cap is hit, the TNSP may then “relax”, knowing that it now has the rest of the week to get the asset back in service.

A TNSP can take various operational actions to manage shortfall costs, depending upon the cause of the shortfall. For example, shortfalls caused by planned outages can be avoided by rescheduling or cancelling the outage; shortfalls caused by forced outages can be managed by expediting the return to service; shortfalls under system normal conditions might be manageable through flowgate support agreements or even by undertaking some additional network expansion.[112]

Each of these actions has an associated cost, which creates a threshold in the TNSP response: if it costs $100,000 to manage a shortfall, the TNSP will only do this if the capped shortfall cost is expected to exceed $100,000. The same action may then be

111 Possibly the TNSP has an incentive to increase them, if benchmarks in subsequent years are based on historical shortfall costs, or, by bringing forward scheduled maintenance, avoid future costs. The AER would need to be careful to avoid introducing such “tanking” incentives into its benchmark-setting process.

112 This is discussed further in section 8.2.3.

taken whether the capped shortfall cost is $100,000, $1m or $10m. So, setting a cap that is much higher than $100,000, say, may be inappropriate:[113] it does not change behaviour and unnecessarily uses up the risk budget.

Correspondingly, each shortfall condition is likely to have an associated timescale. A planned outage is likely to last for weeks but a forced outage will typically last for hours or days. So, the monthly cap may only have a limited impact on incentives to manage forced outages but a large impact on incentives to optimally schedule planned outages. Some numerical examples are presented in appendix C.2 to illustrate these factors.

All of these things should be considered by the AER in setting the scheme. Over time, the AER could learn by experience: if it appears that forced outages, say, are not being managed efficiently, because of a particular cap, the AER might raise this cap when it reviews the scheme.

Settlement

It is proposed that the incentives scheme is settled in the following year. However, a core element of the OFA – proposed in the TFR and adopted by SCER – is that the incentives scheme is settled through access settlement. This significant change to the OFA design requires some explanation.

The TFR approach was proposed for two reasons:

- it was convenient, given that most of the necessary settlement calculations and mechanisms already existed in access settlement and should not need to be duplicated in a separate process; and
- the TNSP contribution to shortfall costs provide generators with a hedge against access firmness and generators would want to monitor these hedge payments – for trading reasons – as close to real-time as possible.

These two issues are explained below and discussed in the context of the revised incentive scheme design.

Shortfall costs are readily calculated in access settlement, since the shortfall amounts and flowgate prices are known for each congested flowgate. Flowgate tags would allow these costs to be allocated to TNSPs.

However, applying the *nested caps* in access settlements is somewhat harder. A running accumulation for each capping period would need to be maintained and adjusted retrospectively as various caps were hit. Allocating the benchmark amount between firm participants is harder still, because none of the required information is available in access settlements. In fact, even if the capped shortfall payments were to be allocated

113 The appropriate level of the cap depends upon the variability of the shortfall cost as well as its expected level. For example, if the shortfall cost had a 1% chance of being $10m but would otherwise be zero, the TNSP is unlikely to take any action if the cap is set at $100,000.

between generators as part of access settlement, a separate process would still be needed to allocate the annual benchmark.

The introduction of the nested caps adds to the complexity of analysing and predicting the incentive scheme payments. As discussed in the previous section, the nested caps allow improved targeting of incentives. On the other hand, they substantially reduce the hedging value of the incentive scheme payments. For example, if a generator is relying on them as a hedge then this hedging effect will suddenly be lost when a nested cap is hit - and then reactivated once the capping period expires.

The incentive scheme is designed to be effective in delivering an efficient level of access firmness, *not* to be an effective hedge. In any case, the potential magnitude of shortfall costs and the relatively limited ability of a TNSP to absorb these on its balance sheet, limits the ability of a TNSP to provide effective hedges.

To conclude, it is neither easy nor desirable to settle the incentives scheme through access settlements. Year-after settlement is simpler and more suitable.

5.3.7 Cross Regional Issues

NEM-wide Concepts

The FAS is conceptually "NEM-wide": ie, applying uniformly across all NEM regions. It is specified in terms of the concepts of shortfall amounts and costs, whose definitions do not explicitly recognise regional boundaries. An analogy is the dispatch process, where AEMO applies the same policies and principles within and between all regions.

The FAS is specified on a flowgate basis. It applies equally to both intra-regional and inter-regional flowgates. Therefore, the firmness of intra-regional and inter-regional firm access should be similar; FAS does not distinguish between them. If, practically, intra-regional firmness is likely to be firmer or less firm than inter-regional firmness, this is due to the physical topology of the power system rather than any bias or discrimination embedded into the FAS.

Regional Specifications

On the other hand, FAS will be *specified* on a regional basis. FAPS conditions will be specified by each TNSP individually, albeit that they must all comply with the NEM-wide FAPS guidelines promulgated by AER. Similarly, incentive schemes will be specific to each TNSP, albeit all designed by the AER with common objectives.

This apparent contradiction between NEM-wide concepts and regional specifications is endemic to the NEM design, being manifested in areas such as reliability standards, TUOS pricing and transmission planning. It reflects the regional nature of TNSPs and jurisdictions, as well as the differing geographical characteristics of the regional networks.

A firm participant may participate in flowgates in multiple regions. That will generally be the case for DICs, for obvious reasons, and also sometimes the case for firm generators. The regional FAS specification means that such participants have a sort of hybrid FAS: a blend of the FAS specifications in the regions in which they participate. It will generally be the case that, even for such participants, the shortfall costs from one region dominate and so the FAS blend will approximate to the FAS in that region.

Cross Regional Settlement

Multi-region participation will give rise to the need for cross-regional settlement processes in relation to both FAPS and FAOS. FAPS-related costs for TNSPs are reflected in the access prices charged to firm participants when they procure access.

Cross-regional settlement of access charges is discussed in chapter 7.

Cross-regional settlement of the incentive scheme was discussed in section 5.2.3.

Transmission Planning and Operation

There are likely to be cross-regional drivers of flowgate capacity. For example, the outage of an asset located in one region may well affect the capacity of the flowgate in another region.[114] Furthermore, some flowgates may be tagged as cross-regional, with multiple TNSPs jointly having responsibility under FAPS and FAOS.

TNSP cooperation and coordination in planning and operation will be needed to address and manage these cross-regional dependencies. Since there are just a handful of TNSPs and they have a common objective of efficiently managing shortfalls and shortfall costs, it is anticipated that TNSPs will voluntarily develop appropriate cross-regional frameworks and processes. Specific Rules and regulations addressing this area may not be required.

It is critical, however, that the allocation of FAS responsibilities between TNSPs is transparent, stable and appropriate. Therefore establishing appropriate governance of the flowgate tagging process will be important.

---

114 According to its flowgate tag.

6 Access Pricing

6.1 Overview

When a TNSP agrees to provide new or additional firm access, this automatically increases the network capacity that the TNSP is required to provide under the FAPS, thus imposing new costs on the TNSP. A fundamental principle of the OFA model is that the firm generator must pay an amount to the TNSP that covers these *incremental costs*. An *access price* is paid by firm generators to TNSPs. The price reflects the incremental transmission costs that are created by the generator's decision to locate in a particular part of the network.

Transmission planning is a long-term process and it is not sufficient to simply calculate the *immediate* cost of the extra expansion required prior to the new access commencing. The new access may cause a *future*, already planned expansion to be *brought forward*. The capital cost remains the same, but the advancement means that, after applying a discount rate, there is an incremental cost in *net present value* (NPV) terms. A methodology in which all incremental costs are calculated – present *and* future – is referred to *here* as long run incremental costing (LRIC). LRIC forms the basis for the access pricing approach.[115]

LRIC is defined to be the difference between two costs: the *baseline cost,* which is the NPV of the *baseline network development scenario* which is in place before the access request is received; and the higher *adjusted cost*, which is the NPV of the *adjusted network development scenario*: an amendment to the baseline plan to accommodate the new access request:

LRIC = adjusted cost – baseline cost

The development plans are derived using a *stylised* methodology which, by assuming away some of the complexity inherent in transmission planning, provides stable and smooth expansion outcomes. The methodology is unlikely to capture every aspect of the network and would involve some judgements about future outcomes, but within these limitations it should be a robust basis for determining access charges.

To ensure that the calculated LRIC is nevertheless realistic and representative of actual expansion costs, critical features that determine LRIC characteristics are included in the methodology. These features include: the measurement of *existing* spare capacity; the *lumpiness* of transmission expansion; the *topology* of the existing transmission system; and the *background growth* of demand and firm generation.

115 Terminology in this area is imprecise and this approach might be referred to as long run marginal cost (LRMC) in other contexts. In this document, LRMC is given a different meaning, so the distinction between LRIC and LRMC is important.

6.2 Design Blueprint

6.2.1 Role of Access Pricing

Overview

Purchasers of firm access are required to pay an access charge which reflects the incremental cost to the TNSP of providing the firm access. Only costs associated with FAPS compliance are included.

Procurement timescales can be separated into the short-term and long-term. In the short-term, flowgate expansion is not possible and so there can be no associated expansion costs. Therefore, access *pricing* – which estimates these incremental costs – applies only for long-term procurement. Access charges for short-term procurement are determined by the auction outcome.[116]

The access price determines the *actual* access charge for long-term *intra*-regional procurement. For long-term *inter*-regional procurement, it sets a *minimum* access charge, through a reserve price in the associated auction. This is discussed further in section 7.2.3.

The access price is predicated on an *access request*, a specified requirement for new access which is defined by the service parameters of registered access:[117]

- amount;
- location; and
- term.

In the context of access pricing, the first year of the term is referred to as the *base year*.

The access price provides a locational signal for generators entering the market and requiring firm access, and for those leaving the market and not requiring firm access. In this context, it is important that the generator knows the access charge prior to committing to its location. This requires the access price to be *fixed*: ie, the charge should be set prior to the generator committing to the purchase and should not be substantially varied subsequently.

In summary, there are three key aspects to an access price:

- it reflects the cost of firm access provision;
- it includes only FAPS-related costs; and

---

116 See section 7.2.4.

117 See section 2.2.4.

- it is fixed for the term of the firm access.

These aspects of access pricing are discussed further in the sections below.

Cost Reflectivity

An objective of OFA is to improve the *coordination* of transmission and generation planning. Currently, TNSPs decide on transmission investments and generators decide on the timing, size and location of new generation. Although information is transferred and shared between these two sectors, each makes its own decision based on its own incentives. A generator will seek to profit maximise, based on estimates of its own revenues and costs. It will not take into account any costs that are borne solely by the TNSP.[118]

Under OFA, an estimate is made of the costs that the generator imposes on a TNSP and this estimated cost is charged to the generator through the access charge. The generator then factors that cost into its decision-making: on where to locate and how much firm access to purchase.

There are two fundamental aspects to this access charge. First, it must reflect the *incremental* cost to the TNSP: the extra cost imposed on the TNSP, not costs previously incurred or future costs that would be incurred irrespective of the generator's decision.

Second, it should reflect the *long-run* cost to the TNSP: the extra costs incurred over the full term of the firm access and beyond.[119] If there is sufficient spare flowgate capacity then the TNSP will incur no *immediate* costs, but the new firm access will deplete this spare capacity and may bring forward the time at which capacity must be expanded.

The costing method used for access pricing, and based on these two requirements, is referred to as a *Long Run Incremental Cost* (LRIC) approach.

Non-FAPS Costs Excluded

Under OFA, a TNSP must comply with three network standards:

- the Firm Access Planning Standard;
- the Firm Access Operating Standard; and
- existing reliability standards.

The issuance of new firm access may create long-run incremental costs associated with all three standards. However, only costs associated with the FAPS are estimated and

---

118 Reliability standards may require a TNSP to expand transmission capacity from that new generator's location to supply increasing demand.

119 In fact, there are liable to be extra *savings* to the TNSP accruing beyond the firm access term, as any transmission asset built to provide the firm access is likely to live beyond the term and help the TNSP to provide firm access into the future.

included in the access price. FAOS and reliability-related costs are excluded, for reasons presented below.

As discussed in section 5.2.3, FAOS-related costs are recovered from generators through the incentive scheme. To include them within the access price would lead to "double-dipping".[120]

The incremental costs associated with the reliability standard (RS) would generally be zero or negative. Any extra transmission expansions that the TNSP will need to make pursuant to FAPS will also help it in maintaining the RS, removing or reducing the need to develop *reliability expansions*[121] associated with RS compliance. The access price is *not* adjusted to reflect these incremental, reliability-related, savings.

FAPS-Related Costs

FAPS-related costs arise from the TNSP obligation, under FAPS, to ensure EFGX is no lower than TFGX. TFGX is the aggregate of target firm entitlements. When a generator purchases firm access, it is provided with target firm entitlements equal to the product of the firm access amount and its participation factor. In the context of access pricing, this amount is referred to as the *incremental usage* associated with the new access. Thus, on every flowgate, TFGX increases by the amount of incremental usage. Of course, on many flowgates, the incremental usage will be zero.

To meet the higher target, the TNSP may need to invest immediately in additional network capacity or a future planned network investment may be brought forward. This increases TNSP costs – in present value terms – in both cases.

The target flowgate capacity is only required under FAPS conditions, so the pricing calculations must be based on assumptions consistent with these conditions.

Non-firm generators have zero firm entitlements and so do not impact on TFGX. Therefore, TNSPs do not bear any FAPS-related costs associated with non-firm generators and, as a consequence, there are no access charges for non-firm generators.

Access Prices are Fixed

As noted, the access price will be based on an estimate of long-run costs: the present value of future costs incurred as a result of the new firm access as well as immediate costs. The timing and amount of future costs will depend upon factors – such as

120 In principle, any elements of operating cost which can be reasonably estimated could be included in the access price. For example, the annual operating and maintenance cost of a transmission asset can often be reasonably estimated using a percentage of its capital value. This cost could be included in the access price through a simple mark-up to the estimated LRIC capital cost. These costs would then be excluded from consideration within the incentive scheme, to prevent double-dipping.

121 A reliability expansion is one that is *only* required in relation to RS compliance: ie, FAPS would be maintained in the absence of the expansion.

demand growth and future firm access purchases – which must be projected. Actual costs will turn out differently to estimated cost if the projections are wrong.

Of course, projections are almost *always* wrong and so actual costs will almost inevitably differ from estimated. This may lead to *regret*, where decisions made on the basis of the estimated costs (embedded in the access price) would have been different had the actual costs be known. But this is, unfortunately, a feature of all long-run decision making. The generator entering the market will similarly do so on the basis of its own estimates of future costs and revenues, which must also be based on projections.

There are two consequences of this. First, it is not possible or even meaningful for the access prices to be *precise*, in the sense that they will be very close to actual costs. Forecast errors mean that actual costs will almost always diverge from estimated. An alternative objective is that prices are *unbiased*: that is, they are an accurate estimate of *expected* costs.[122] However expected costs are never revealed, only actual costs.[123] So assessing empirically whether prices are unbiased is problematic.

Forecasting uncertainty and errors will typically increase the further ahead the forecast is made. Thus, estimates of near-term costs are likely to be less affected by forecasting uncertainty, and so potentially more accurate, than longer-term costs.

Second, because access charges are *fixed*, any variance between actual and estimated costs will be borne by the TNSP initially and by TUOS customers to the extent that TUOS regulation permits these variances to be passed on through adjustment to TUOS prices.[124]

Given the inherent difficulty of accurately estimating incremental costs, it is important to avoid the mistake of striving for *spurious* precision: that is to say, introducing additional complexity into access pricing that would improve precision in the hypothetical situation where the future is known but, given forecasting uncertainty, does nothing to improve actual accuracy or level of bias in the estimates.

Pricing Model

Access prices will be calculated by a *pricing model*. The model contains forecasts of market and transmission conditions. When a particular access request is input into the pricing model, a dollar access price is determined.

As noted above, the FAPS requires that a TNSP maintains and expands flowgate capacity so that EFGX equals or exceeds TFGX under FAPS conditions. So the pricing

---

122 Mathematically, this means the *average* of the actual cost, assessed over all possible future scenarios and weighted by the probability of those scenarios eventuating.

123 In practice, it is difficult even to define and measure actual incremental costs, since only a TNSP's *total* costs are known and it is difficult to attribute these costs between different firm generators and customers.

124 This is discussed further in section 8.2.2.

model estimates the costs arising on each flowgate, associated with this FAPS requirement.

As previously discussed, flowgates fall into two categories: thermal flowgates and stability flowgates. These different flowgate types arise at different locations and have capacities driven by different factors. Therefore, two different pricing models are used for estimating the cost of maintaining TFGX on the two flowgate types.

The thermal pricing model is based on:

- a simulated firm generation dispatch under FAPS conditions; and
- a stylised expansion plan that is designed to ensure that this dispatch remains feasible as the market grows.

The following two sections discuss these two aspects of thermal pricing. The final section considers access pricing for stability flowgates.

6.2.2 Thermal Load Flow Model

Overview

As discussed in section 5.2.2, the FAPS requirement is similar to the requirement that, under FAPS conditions, the transmission network is able to accommodate a *firm generation dispatch*: ie, a dispatch of all firm generators and DICs at their target firm access level. The two requirements are equivalent in the special case where the output of all flowgate support generators, under FAPS conditions, equals their target firm access levels. To simplify the modelling of access costs, this assumption is implicitly made in relation to thermal flowgates and so the pricing model can be based on the dispatch analogy.

Recall that target firm access is the lower of the registered access and generator capacity. In the context of access pricing, generator capacity is *ignored*.[125] Therefore, the target firm access level is assumed to be equal to the registered access level.

Load Flows

For thermal flowgates, the adequacy of flowgate capacity against FAPS is assessed by using the dispatch analogy: ie, testing that the firm generation dispatch is feasible. Feasibility is tested using a simplified *DC load flow* model of the network, which estimates the power flow on every *network element* (ie, line or network transformer) under the firm generation dispatch. The assumed levels of local demand and non-scheduled generation are in accordance with the FAPS conditions. As discussed in section 5.3.2, some notional demand or generation is added at the RRN to ensure that demand and generation are in balance across the network.

There are no transmission constraints placed on this load flow. Where the modelled flow exceeds the thermal rating[126] of the element on one or more elements, the dispatch is infeasible and so the FAPS is not met. In this case, flowgate capacity must be increased on the problematic elements. The modelling of such expansions is discussed in section 6.2.3.

Baseline Forecast

As described in section 5.2.2, the FAPS conditions will describe an annual *FAPS snapshot*. This snapshot will be defined by the various parameters including:

1. forecast demand at each network node;

2. forecast non-scheduled generation; and

3. forecast scheduled and semi-scheduled generation, where this affects flowgate capacity.

The pricing model uses the FAPS snapshots for demand and non-scheduled generation. However, the FAPS snapshots for scheduled and semi-scheduled generators are not relevant for pricing, because the model uses a firm generation dispatch: ie, every scheduled and semi-scheduled generator is assumed to be dispatched at its registered access level. So, forecasts of *registered access*, rather than dispatched output, are required.

Demand forecasts will be based on forecasts developed under existing processes: for example TNSP annual planning reporting. Forecasts for non-scheduled generation will be developed similarly: eg, based on AEMO's National Transmission Network Development Plan (NTNDP).

Forecasts for registered access will be built up from four components:

- *existing* registered access;

- *modelled* access: expected future firm access purchases;

- *reliability access*: where firm access is otherwise insufficient to maintain reliability standards; and

- forecast *renewals* of existing registered access.

Anticipated access requests are *not* included in the baseline. These are discussed further in section 6.3.4.

Forecasting processes for the four listed components are discussed in turn below.

125 This is to avoid the problem of a small generator being able to buy access at a lower price and then subsequently transferring this access to a larger generator.

126 Again, based on FAPS conditions.

Existing registered access is obtained directly from the firm access register.

The forecast for *modelled access* is based on two factors:

- the expected location, timing and size of new generator entry; and
- the expected firmness of new generators.

Generator entry forecasts are provided in the NTNDP and the baseline forecasts could be based on this or developed using a similar process. The expected firmness could be based on the firmness of similar, existing generators.

The concept of *reliability access* (RA) is introduced and discussed in section 5.3.5. RA is added to the baseline forecast when total forecast registered access, plus non-scheduled generation, would otherwise be lower than aggregate forecast demand. That reflects the likelihood that RA would be provided – possibly through reliability expansions – under such circumstances. The RA would be allocated to locations where reliability access might be provided. This would be where existing or modelled non-firm or part-firm generators are located.

Renewal means the procurement of firm access that matches some existing, expiring firm access. For example, if there is some existing access at node X with amount Q and a term ending 2020, then it can be renewed through the procurement of firm access commencing 2021, also at node X with amount Q. Existing firm access may, in some cases, be awarded with a *renewal right*, which simply means that its renewal is incorporated into the baseline forecast. The significance of renewal rights is explained in section 6.2.3.

There is no requirement to forecast network capacity, since this is built up cumulatively by adding the modelled network expansions to the (known) network capacity in the base year. The base year will be set far enough in advance for all committed actual network expansions to be in place.[127]

Study Period and Forecasting Horizon

The *study period* is the future period over which the pricing model operates. How far out should this period go?

There are two, conflicting considerations. On the one hand, the expected accuracy of the baseline forecasts will deteriorate the further out the forecasts go. On the other hand, the long-term life and lumpiness of transmission assets mean that, even after applying the discount rate, expansions in the distant future can still have a material impact on the expansion cost and so need to be modelled.

The pricing model resolves this conflict by introducing a *forecasting horizon* within the study period. The load flow model is used, and so complete baseline forecasts are required, up to the *forecasting horizon*. Element flows are then extrapolated beyond this

127 This is discussed further in section 7.2.1.

horizon based on a simple assumption (eg, that the annual increase is constant) that does not require baseline forecasts. With this extrapolation, stylised expansions can be determined for the full study period.

Network Topology

In addition to the baseline forecast, the load flow model requires assumptions on *network topology*: the structure and size of the transmission network. The topology is defined in a database of network nodes and network elements that is used by the pricing model. For each network element, the database contains the following parameters:

- *connectivity*: the names of the two nodes which the element connects;
- *admittance*: the electrical admittance[128] of the element;
- *rating*: the thermal limit;
- *type*: the element type, which determines the stylised expansion; and
- *end-of-life*.

The demand and generation in the baseline forecasts are mapped to the nodes in the network topology.

The network topology changes from year-to-year. In the base year, it represents the existing transmission network.[129] When an asset reaches the end of its life it will be *removed* from the network database. When a stylised expansion is prompted it is *added* to the network topology.

The end-of-life parameter is needed in order to include *replacement* costs in the access price: costs associated with *replacing* network capacity that has been lost due to a network asset reaching the end of its life, as opposed to *expanding* network capacity. In a low-growth environment, replacement costs for a TNSP, overall, can exceed expansion costs. Therefore, it is important these are included, where appropriate, in access prices. Replacement scenarios are presented and discussed further in appendix D.2.

System Security

Flowgates arise from the constraint equations in dispatch. The Rules require that dispatch is *secure*, which means that there must be no thermal overloads or instability on the network:

- *pre-contingency*: under the current power system conditions; and

128 Admittance in an AC network is analogous to conductivity in a DC electrical network or the size of a pipe in a pipeline network.

- *post-contingency*: immediately following any credible contingency.

A credible contingency is any outage of generation or transmission assets following a single failure or trip.

In relation to thermal flowgates, this security requirement means that, for each network element, there are multiple flowgates:

- *one pre-contingency flowgate*: relating to the dispatch constraint that prevents pre-contingency thermal overload; and

- *one post-contingency flowgate for each credible contingency*: relating to the dispatch constraint that prevents thermal overload following that *credible contingency*.

Generally, a different thermal rating will be used for pre-contingency and post-contingency flowgates. Because post-contingency conditions only last for a short period of time, after which generation re-dispatch can remove overloads, an *emergency* or *short-term rating* is used. Pre-contingent conditions last for unlimited periods and so a *continuous rating* is used.

The pricing model considers only those credible contingencies relating to outages of single network elements. This, nevertheless, substantially increases the number of flowgates that must be considered: ie, if there are N network elements then there are N pre-contingency flowgates and N x (N-1) possible post-contingency flowgates. A possible simplification to address this complexity is discussed in section 6.3.1.

6.2.3 Thermal Expansion Model

Overview

The previous section explained how a load flow model can be used to calculate the flow on each network element under a firm generation dispatch and compare this to its thermal rating to test whether the dispatch is feasible. This can be done for every year in a sequence, commencing in the base year, based on each year's baseline forecast and network topology.

When an overload is detected, where the flow exceeds the thermal rating, the pricing model defines a *stylised development* to the network which removes the overload. This expansion is then added to the network topology and so is reflected in the load flow modelling in that year and in subsequent years. The complete set of these stylised developments is referred to as the *baseline network development scenario*.

As discussed above, access prices are based on the long run *incremental* cost. To estimate this, an *adjusted network development scenario* is also calculated, based on a modified *adjusted forecast*, but otherwise using exactly the same load flow and expansion modelling as the baseline calculation. The forecast is adjusted from the

129 Together with any committed expansions that would be in place in the base year.

baseline by adding in the access request to the registered access assumptions. Everything else remains unchanged. By construction, any differences between the two expansion plans must be caused by the access request, *ceteris paribus*.

The pricing model contains assumed costs for every possible stylised expansion. Using these, the model calculates the respective costs (in present value terms) of the baseline and adjusted network development plans, referred to as the *baseline cost* and *adjusted cost,* respectively.

The LRIC is then defined to be the difference between these two costs:

LRIC = adjusted cost – baseline cost

The next sections discuss each aspect of this process in turn:

- the definition of the stylised expansions and calculation of the baseline scenario;
- the definition of the adjusted scenario; and
- the calculation and allocation of present costs.

Stylised Developments

In its *transmission planning* process, a TNSP could address a thermal flowgate shortfall in a number of ways:

- by increasing the network capacity of the relevant network element: this increases flowgate capacity by a corresponding amount;
- by increasing flowgate support: eg, by entering into a flowgate support agreement that ensures that the flowgate support generator will be operating under FAPS conditions; or
- by adding new transmission lines elsewhere or even, in some cases, by removing existing (upstream) lines. This has the effect of diverting some of the power flow onto other network elements.

The variety of options available to the TNSP makes transmission planning complicated, but this is mitigated by the fact that the preferred option only needs to be selected and developed when there is an *imminent* FAPS shortfall: a TNSP does not have to consider possible solutions to potential shortfalls many years in the future, say.

For access pricing, on the other hand, flowgate expansions must be decided for the entire study period. This makes the development of the scenarios far more complex than TNSP planning. On the other hand, given the likely inaccuracy of the baseline forecasts in outer years, and also the discounting factor, it is less important to identify and select the *optimal* expansion. A simple, approximate, development rule is appropriate.

The rule applied in the model is the simplest possible: if there is an overload on a thermal element, a new circuit is added in parallel: ie, between the same two nodes. This new circuit is modelled as being incorporated into the existing element. The thermal rating of the element is increased and so the overload is removed.[130]

The expansion size (ie, its rating) determines the increase in the element capacity. Potentially, multiple expansions may be required on a flowgate in a particular year, where the capacity shortfall exceeds the expansion size.

Adjusted Scenario and Renewal Rights

As explained above, the adjusted scenario is simply the baseline scenario to which the access request has been added. This addition is straightforward in relation to requests for new firm access. However, the case of *renewal requests* is more complex.

In section 6.2.2 it was noted that renewals would be included in the baseline forecast where the associated existing firm access has a *renewal right*.[131] For example, if some registered access expires in 2020 but was forecast, in the baseline, to be renewed until 2030, then the baseline forecast would extend the term of the registered access until 2030.

A *renewal request* is an access request that seeks to replace some expiring firm access that is attached to a renewal right. So, by definition, renewal requests are anticipated in the baseline forecast. The renewal request may not be exactly the same as in the baseline forecast: it could be for a different term or different amount. For simplicity of exposition, it is assumed that the two *are* the same.

When pricing a renewal request, the associated renewal must be *removed* from the baseline, since otherwise the renewal will be double-counted in the adjusted scenario. That would be unrealistic and lead to poor pricing outcomes. So, the renewal is *removed* when calculating the baseline cost and then *replaced* when calculating the adjusted cost. So, in this case, the access price is the difference in cost between the following two scenarios:

- the baseline scenario; and
- the baseline scenario with the renewal request removed.

In this case, the access price is the incremental *saving* to the TNSP associated with *removing* the renewal from the baseline. This is referred to as the Long Run Decremental Cost (LRDC).

---

130 The increased admittance of the expanded element may actually cause flows to increase and so, possibly, the overload is not removed. However, for simplicity, this effect is not modelled; the admittance is assumed to be unaffected by the expansion. This is discussed further in 6.2.4.

131 The renewal right means only that the renewal is included in the baseline. It does not confer any other rights associated with access procurement.

Thus a renewal request is priced at LRDC. Other access requests, including those that relate to the renewal of firm access that does *not* have renewal rights, are charged LRIC. Since the LRDC will typically (but not always) be lower than the LRIC, the renewal right does confer some financial benefit to the purchaser.

Renewal rights are considered problematic for a number of reasons and so are not awarded in relation to *purchased* firm access. This is discussed further in section 6.3.4. However, renewal rights are proposed to be awarded on transitional access, as discussed in section 9.2.2. Consequently, only TA renewals are included in the baseline forecast.

Calculating the LRIC

As discussed, the LRIC is based on the difference between the present cost of the adjusted network development plan and the present cost of the baseline network development plan. This is a simple, dollar amount. It is calculated in several stages.

First, the exact *timing* of each stylised expansion must be defined. As discussed above, an expansion must occur when a capacity shortfall arises in a particular year. The *date* of the expansion within that year is defined as the point in the year when the spare capacity would be exhausted if demand and generation grew linearly (say) through the year, from one year's FAPS snapshot to the next. Although it would be simpler – and perhaps more representative of actual TNSP planning - to just specify all expansions as occurring on the 1st July (say), this could lead to discontinuities in the pricing curve when a small change in the request access amount advances a large expansion by a full year in the adjusted expansion plan.

Second, a *discount rate* must be defined. This would be based on the *regulated WACC* of the relevant TNSP. The present cost of each individual expansion would be discounted to the current time by applying this discount rate.

Third, the present costs of the two expansion plans – and so the LRIC – are calculated *separately* for each element. For some elements, the adjusted cost will be less than the baseline cost. In this case, the element LRIC is specified to be zero: ie, the element LRIC is not permitted to be negative. The reasons for this are discussed in section 6.3.4.

Finally, the costs must be allocated between TNSPs. Although the generator submits the access request to its local TNSP, it is possible in some cases that some elements on which a positive LRIC has been calculated lie in a neighbouring region. The associated expansion costs will be borne by the neighbouring TNSP. Therefore, the element LRICs will be aggregated by region to give an access price that is divided into regional amounts.

6.2.4 Simplifications

Overview

The pricing model simulates transmission planning under FAPS in order to estimate future costs under the adjusted and baseline scenarios. Despite substantially simplifying this process through the use of stylised expansions, the modelling requirements remain dauntingly complex.

A number of further simplifications are proposed. These have the effect of removing the complex interactions between different network elements and their associated expansions. This simplifies model development and, more importantly, it allows pricing outcomes to be more transparent. It means that *separate* and *independent* expansion plans and costs are determined for each element. Typically, for a particular access request, the costs arising on a dozen or so elements contribute most of the overall price. Further analysis of these elements will provide understanding of, and insights into how the access price comes about and how it depends upon the request parameters. It is not necessary to analyse every single element.

The simplifications that are made are:

- use of a lossless DC load flow in the load flow model;
- fixed participation factors;
- fixed security adjustment; and
- correction for meshedness.

These are discussed in turn below.

DC Load Flow

A DC lossless load flow simplifies the calculation of load flows on a power system by ignoring reactive power and electrical resistance. It allows the flows to be expressed as a linear combination of the nodal generation and demand quantities.

For any specified network topology, a set of participation factors can be calculated. These are similar to the participation factors appearing in flowgate formulations, although they apply to demand as well as generation. The flow on any network element is then just a linear combination of the generation (ie, registered access) and demand forecasts:

$$\text{Flow on element} = \alpha_1 \text{ x } (G_1 - D_1) + \alpha_2 \text{ x } (G_2 - D_2) + \ldots \alpha_N \text{ x } (G_N - D_N)$$

Where:

α_i is the participation factor for node i for the particular element

D_i is the local demand at node i

G_i is the generation at node i

Fixed Participation Factors

Because the participation factors depend upon network topology, they will change when the topology changes: ie, when there is an expansion, or when an element reaches the end of its life. The model can be simplified by *ignoring* these changes. The participation factors are calculated only for the *base year* network topology and then left unchanged for every year up to the forecasting horizon (beyond which they are not used).

Fixed Security Adjustment

Section 6.2.2 discussed the need to consider post-contingency flowgates, which reflect the constraints that would be included in a secure dispatch. If we define the *spare capacity* of an element as the difference between the rating and the flow, the post-contingent spare capacity may differ from the pre-contingent level due to two factors:

- the post-contingency thermal rating is higher, because an emergency rating is typically used; and
- the outage will cause the power flowing, pre-contingency, on the failed element to be redistributed across the remaining elements.

For the dispatch to be feasible and secure there must be spare capacity under all conditions: ie, it is the *minimum* spare capacity (across all conditions) that determines the need for expansion.

The *security adjustment* is therefore defined according to the formula:

> Security adjustment = pre-contingency spare capacity – minimum spare capacity

By itself, this does not offer any simplification, since every contingency must still be analysed in order to determine the security adjustment. Simplicity is introduced by calculating the security adjustments for the *base year* only and then *assuming* they remain unchanged in subsequent years. The minimum spare capacity is then able to be calculated without having to consider all of the post-contingency conditions: ie, minimum spare capacity = pre-contingency spare capacity – fixed security adjustment.

Meshedness

The fixing of participation factors creates an increase in the effective *lumpiness* of the network. This is explained in a simple example below.

Consider three identical elements in parallel. Each will have the same rating and flows and so will require expansion in the same study year. In reality, if one were expanded, this would divert flows from the other two meaning that their expansions could be postponed. This would be reflected in lower participation factors in the DC load flow.

However, with the participation factors fixed, there is no flow diversion and so all three elements would be expanded in the same year. This would increase spare capacity across the three elements far more than would be needed in practice and the associated present expansion costs would also be higher than necessary. For example, if the stylised expansion were sized at 500MW, an expansion of 1500MW (3 x 500MW) would be triggered when only 500MW were really needed. So, the pricing model sees an *effective lumpiness* of 1500MW. The capacity across the three elements is expanded in 1500MW steps rather than in 500MW steps.

To correct for this bias, the size of a stylised expansion is reduced by a *meshedness factor* that describes how many transmission paths there are in parallel to the element. A radial element has no (other) parallel paths and so has a meshedness factor of one. In the example of the three elements the meshedness is at least 3: possibly higher if there are other parallel paths. So, in the example, the total modelled expansion would be divided by three and so would be similar to what it would be if the change in admittance was modelled.

Meshedness is defined formally in appendix D.1.

Element Expansion Scenario

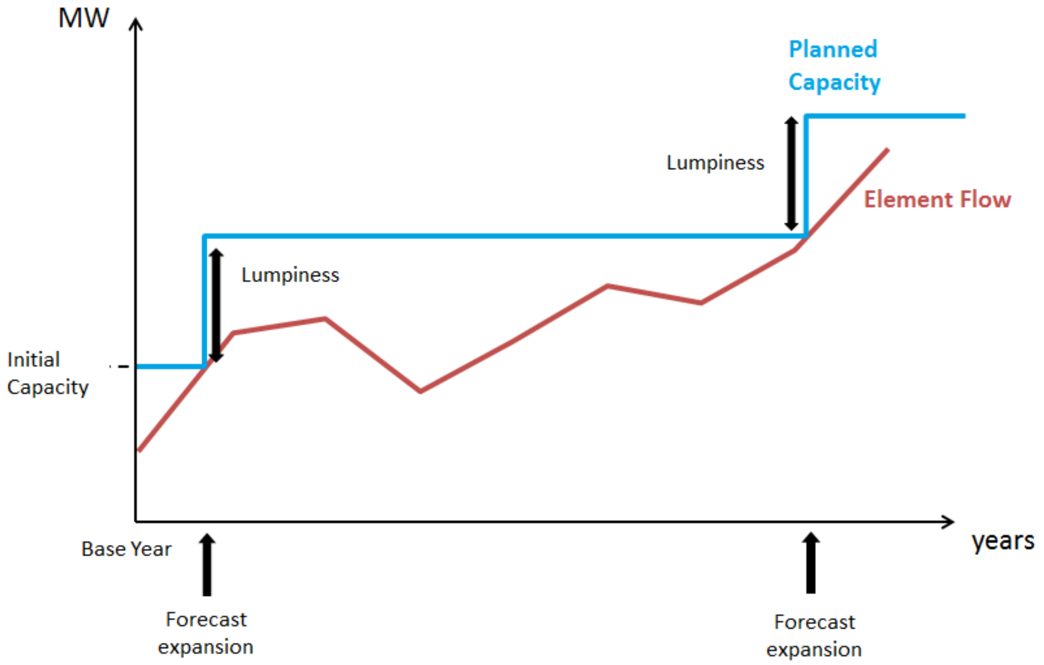
In a *real* planning process, there are several ways in which elements interact, meaning that the expansion of an element cannot be planned in isolation: the expansion of other elements must be taken into account. The expansion of an element will depend upon:

- the losses on other, upstream elements, reducing the load flow;
- the diversion of power through other elements on parallel paths: ie, affecting the participation factors; and
- the contingent outage of another element in the security assessment: ie, affecting the security adjustment.

The three simplifications discussed above – together with the simplification embedded in the stylisation of expansions – remove these interdependencies for all years but the base year. This allows expansion scenarios on each element to be calculated and analysed independently.

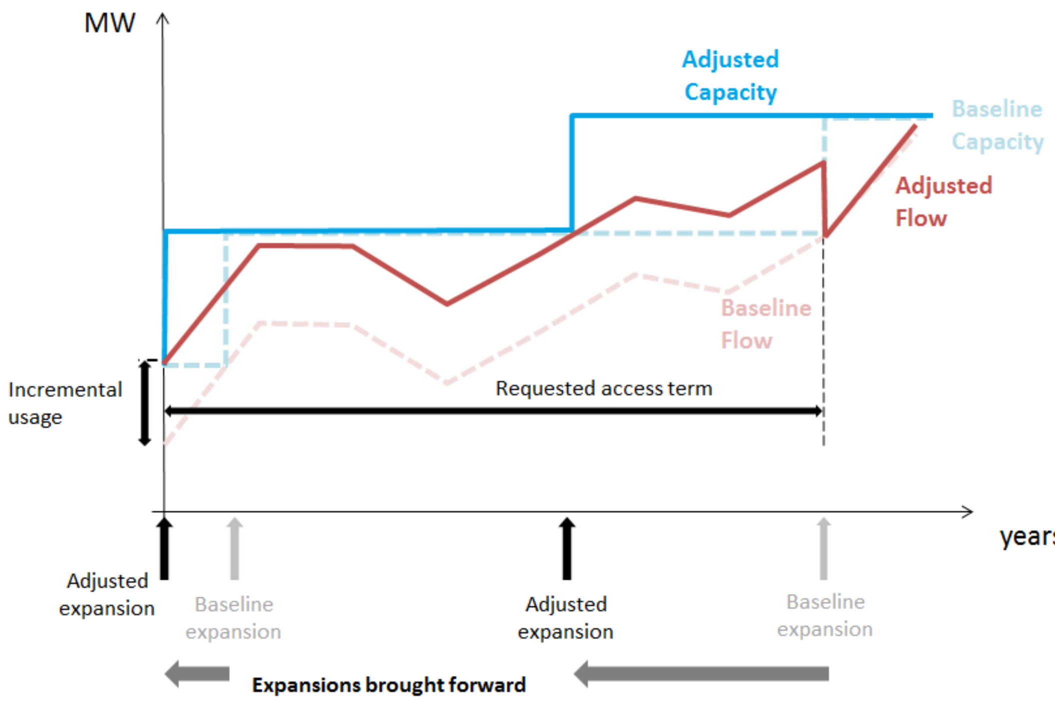
A typical baseline expansion scenario for an element is presented in Figure 6.1. The vertical axis shows element capacity (ie, rating) and flow. These evolve over the study period. Flows change in response to corresponding changes in the baseline forecast. Capacity changes as expansions are prompted in response to the flow growth.

Figure 6.1 Baseline Network Development Scenario



The adjusted expansion scenario is presented in Figure 6.2. The adjusted flow scenario is higher than the baseline flow, increasing by the incremental usage over the term of the access request.

Figure 6.2 Adjusted Network Development Scenario



In the example, the higher flow prompts an *immediate* expansion in the adjusted expansion plan as the initial spare capacity is exhausted. Subsequent adjusted expansions are *advanced* compared to the baseline plan.

The graphs demonstrate how the complexities and intricacies of transmission planning have been distilled down to their simplest possible form, through the use of stylised expansions and the three simplifications discussed above. At this level, access pricing appears simple and intuitive. However, behind the scenes, the complexities of network topology and baseline forecasting, of course, remain.

6.2.5 Stability Flowgates

Overview

The model described above calculates an LRIC reflecting the cost of expanding *thermal* flowgates. The access request may also create incremental costs for a TNSP in relation to expanding *stability* flowgates. This will occur where the relevant generator participates in a stability flowgate and so prompts either an immediate expansion or advances baseline expansions of that flowgate.

The thermal LRIC model uses a DC load flow to calculate spare capacity on elements under a firm generation dispatch. Such a model cannot model stability flowgates. A separate pricing model is needed for estimating costs associated with stability flowgate expansion.

It is currently proposed that:

- a *rule-of-thumb* approach is used for pricing of intra-regional access; and
- a *Deep Connection Charge (DCC)* approach is used for pricing of inter-regional access.

These two approaches are outlined below. The DCC approach is considered in more detail, in the context of thermal pricing, in section 6.3.5. The reasons for this proposed approach to stability pricing are discussed further in section 6.3.6.

In each case, the calculated stability access price is added to the thermal access price to give an overall access price upon which access charges are based.

Intra-regional Pricing

Under a rule-of-thumb approach, a simple $/MW price is calculated at each generator node and the access charge is then simply the product of this price and the requested amount. The $/MW prices could be based on an analysis of a sample of expansions: ie, a TNSP would undertake detailed planning of transmission expansion to accommodate a number of representative access requests and then use statistical analysis to identify prices which would give reasonable estimates of access costs for all other possible access requests.

Such an approach may be adequate where stability costs are relatively low compared to thermal costs, albeit not so low that they can be ignored completely. In this context, the significant approximations implicit in the rule-of-thumb methodology would not adversely affect the precision of access prices overall.

Inter-regional Pricing

Unlike LRIC, a DCC pricing approach only estimates the immediate expansion costs associated with an access request: ie, the expansions that are needed prior to the commencement of the requested access, to ensure that the network is FAPS compliant in the base year.

DCC is inferior to LRIC in that it ignores future costs: ie, the advancement of baseline expansions in the adjusted expansion plan. It is for this reason that an LRIC approach is preferred for estimating thermal expansion costs.

On the other hand, DCC is a fundamentally simpler approach, since it only requires base year expansions to be identified and costed. In this respect, it is similar to the TNSP planning process and it becomes feasible to base access pricing on actual TNSP plans rather than the stylised plans used in the LRIC pricing model. This makes DCC a more tractable approach for stability costing.

Under a DCC pricing methodology, a TNSP would undertake detailed planning to identify the base year expansion requirements. The identified expansion is likely to consist of:

- thermal assets: ie, new transmission lines and transformers; and
- stability assets: stability equipment such as capacitors and SVCs.

The total cost of these new assets would be calculated and incorporated into the access price. In addition, the new thermal assets would be *hard-coded* into the LRIC model used for costing thermal expansion. This means that the transmission conditions in the base year for the *adjusted* scenario would now reflect the new thermal assets as well as the existing transmission network.[132]

Hard-coding is discussed further in section 6.3.2, which explains how the thermal expansions are still priced using LRIC, despite being hard-coded.

132 The new stability assets would not generally affect thermal elements in a DC load flow and so do not need to be hard-coded.

6.3 Design Issues and Options

6.3.1 Pricing Objectives

Overview

The access pricing model has been developed with three objectives in mind:

- robustness;
- transparency; and
- automation.

These three objectives, and the reasons for having them, are explained below.

Robustness

Section 6.2.1 discussed the conceptual difficulty of defining the *precision* of a pricing model, when prices are predicated on long-term baseline forecasts that will inevitably turn out to be incorrect. Rather the objective of pricing should be that access prices are *unbiased*: the price reflects the expected LRIC, taking into account the range and probabilities of future scenarios.

Does every price quoted need to be unbiased? Or is it sufficient that quoted prices are, collectively unbiased, so that any individual biases average out over time and that there is no bias to the biasedness? After all, the overall impact on TNSPs depends on *aggregate* access revenue compared to *aggregate* access costs.

The problem is that firm access is *optional*. Generators are more likely to purchase access when it is priced too cheaply, and less likely to purchase it when it is priced too high. Generators will tend to "cherry pick" the low prices. This may lead to access revenue not covering access costs in aggregate, even if the biases average out.

Robustness of pricing, in this context, means the level of certainty that quoted prices would not be *too* biased – won't be too far from the expected cost, up *or* down. Small variations do not matter too much: they won't give rise to cherry picking and so they will average out over time. Large variations would not average out.

Robustness is easier to monitor than accuracy or bias, since gross errors may well be obvious. Ideally, such gross errors would be removed before the prices were formalised into offers. However, such a process would conflict with the automation objective, discussed below.

Transparency

Access prices are intended to provide signals to generators, guiding decisions on location and firmness. If prices lack transparency, it will be hard for generators to follow the signals: they would have to obtain indicative prices for every possible combination of location and firmness level and compare them all. As discussed, these prices can then change as the baseline changes, putting the generator back at square one.

If prices are transparent, generators will understand how they relate to factors such as distance from the RRN, connection voltage, current and forecast congestion and so on. Simplicity can contribute to transparency. So does *continuity*: meaning that small changes in the parameters of the access request give rise to only small changes in prices.

Automation

If the pricing model is *automated*, it can be operated by anyone. The model will need to be designed and populated with the baseline forecasts and network topology. But, once this is done, any person can, in principle, input an access request, run the model, and obtain an access price. If pricing involves, on the other hand, expert judgement, in relation to *each* access request, then automation is no longer possible.

There are several advantages of an automated model:

- *speed*: a pricing application can easily and quickly be run on a computer;
- *transparency*: even a model that is not conceptually transparent (eg, a "black box") can become empirically transparent if a user can run it enough times to infer a pricing pattern; and
- *governance*: with an automated model, the entity responsible for the model[133] can be remote from the procurement process.

Implications

It is difficult to achieve all three objectives. As discussed, the pricing model has been developed with several simplifications and stylisations that should improve transparency. Such a model might not be robust; there are likely to be some situations where the stylised expansion is a poor approximation to the actual expansion the TNSP would undertake.

The proposed pricing model, and associated governance, reflects a prioritisation of the transparency and automation objectives. Possibly, as a result, the pricing model is insufficiently robust. Some possible changes to the design of the pricing method to

133 Which is proposed to be the AER.

improve robustness are discussed in the next section. These will inevitably degrade automation and/or transparency.

6.3.2 Improving Robustness

Overview

As discussed above, the robustness objective is sacrificed, at least to some extent, by the emphasis on transparency and automation. This section considers some ways in which access pricing might be modified if the objectives were rebalanced; they all involve introduce some further complexity or expert judgement in order to improve pricing robustness.

Raising Disputes

The access pricing model involves myriad assumptions and estimates: on baseline forecasts, expansion costs and so on. The AER, who will have responsibility for populating the model with this information, will endeavour to make these inputs as fair and accurate as possible, but pricing outcomes may reveal possible errors or inconsistencies in input assumptions. These might be identified by the generator or by the TNSP. The generator might see price variations between different requested locations or quantities that appear arbitrary or counter-intuitive. The TNSP may have a prior understanding of the likely expansion costs and consider that the access price substantially underestimates or overestimates them.[134]

A mechanism could be established that would permit any party to dispute an access price outcome or an associated input assumption. This would go beyond the usual dispute process: ie, the dispute would not assert that there had been a breach of the Rules, but simply that the price or assumption appeared unreasonable and should be reviewed.

Such disputes will obviously slow down the process of access pricing, with possible delays for those behind in the procurement queue. Possibly disputes could be raised only in relation to *indicative* prices to avoid this. Or, perhaps, raised even earlier, when the model assumptions are first published.

Menu of Stylised Expansions

The stylised expansions used in the thermal expansion model are inherently simple. Possibly, a *menu* of alternative expansions could be introduced and some automatic logic of choosing between them included in the model. Perhaps, even, the TNSP could use expert judgement to choose between them, at least for near-term expansions.

---

134 Intuitively, one might think that a TNSP will only be concerned if the access price is too *low*. But high prices may cause TNSP concern as well. A high price may deter generators from purchasing firm access and so deny a TNSP the opportunity to (profitably) expand its network.

Hard-Coding

The logical extension of an expansion menu is for a TNSP to design *bespoke* expansions, especially any *immediate* expansions, using its normal planning tools. As noted,[135] this may be necessary anyway in relation to stability pricing of inter-regional access.

Hard-coding is not inconsistent with an LRIC approach. The designed expansion would be hard-coded into the adjusted expansion plan by making the necessary changes to the network topology. This adjusted network topology would be used as the starting point for the adjusted expansion calculation, whereas the usual "existing plus committed" topology would be retained for the baseline calculation. The model would then be run as usual to determine baseline and adjusted expansion scenarios. The adjusted scenario would include the hard-coded, bespoke immediate expansions, followed by later stylised expansions.

Just as with the stylised expansions, the spare capacity provided by the hard-coded expansion would be implicitly valued by the LRIC model and that value deducted from the expansion cost.[136]

Hard-coding means that access pricing is no longer automatic. It would only be introduced if automation were no longer considered a pricing objective, or if it was absolutely critical to achieving robust pricing.

Hard-coding would inevitably mean some loss of transparency, since generators would not have the expertise to anticipate what bespoke expansions would be required. There would also be some loss of continuity: requests that prompted some immediate, bespoke expansions could have a quite different price to similar requests that did not prompt immediate expansion and consisted only of stylised expansions.

6.3.3 Central Planning and Pricing

Overview

A concern that has been expressed around the role and method of access pricing is that it means that *centralised planning* in the NEM is maintained, or perhaps even further entrenched. There are many aspects to this concern which can usefully be unpacked and analysed:

1. the precise meaning of "central planning" and the reasons why it is a concern;
2. the degree of central planning in the NEM currently; and
3. the impact that OFA, and access pricing in particular, has on the degree of central planning.

135 In section 6.2.5.

136 The concept of valuing spare capacity is discussed in section 6.3.5.

These aspects are discussed in turn below.

Central Planning

A *central planner* is usually defined as an entity that has responsibility for production decisions across an industry sector. By statute or regulation, it is required to make these decisions in accordance with some specified objectives, rather than because the decisions are necessarily in its own best interests: eg, profit maximisation.

In contrast, under *decentralised planning*, disparate producers make individual production decisions, guided by price signals: the cost of production capacity and inputs, and the sales price of the produced goods. Each individual producer aims to maximise profit. There is no *explicit* coordination or cooperation between the producers and with consumers. It takes place *implicitly* through the price mechanism.

Concerns around central planning are based on *information*, *expertise* and *agency*. A central planner is tasked with a complex optimisation problem which relies on accurate information (eg, on current and future demand for the produced goods) and expertise in decision making. Since it is not making decisions in accordance with its own interests, it may be affected by conflicts or take undue risks that others bear the consequences of.

On the other hand, decentralised planning is only suitable where the overall planning objective is consistent with individual profit maximisation and where efficient prices exist to guide and implicitly coordinate the individual decisions.

Current Planning in the NEM

The OFA model is concerned with transmission and generation planning. Under current arrangements, transmission planning is based on a central planning model: inevitably in some ways, because there is just a single responsible planning TNSP in each region. Generation planning is decentralised. Taken together, then, transmission-generation planning is a hybrid of the two models.

An existing concern, that OFA aims to address, is that decentralised planning of generation currently cannot be efficient because some price signals are *missing*. Efficient pricing is needed in a decentralised model to promote co-optimisation. In the absence of such price signals, co-optimisation of generation and transmission planning cannot occur.

Impact of OFA

It should first be noted that OFA does not change the planning model operating in the NEM, at the conceptual level described here. Transmission planning continues to be *centralised*; generation planning continues to be *decentralised*. In that sense, it is illogical to be critical of OFA for maintaining the current degree of central planning, since this is in fact the intention.

The change that OFA introduces is to complete the price signals for generation planning by introducing a price for transmission access. Currently access is provided to generators for free, through TNSPs' application of the reliability standards. The quality of access – and the level of congestion costs – may vary by location and this is a price signal of sorts. However, it is neither complete nor predictable and so is inadequate for facilitating a full co-optimisation of generation and transmission.

Further, while there are still central forecasts made in the OFA model, the generator has to make a decision as to whether to buy firm access or not. This shifts it to a more decentralised planning environment.

Access pricing aims to be efficient by reflecting the expected cost of firm access provision. If it successful in that aim, then the co-optimisation of transmission and generation investment should improve as a result.

Central Pricing

Under OFA, access pricing is undertaken by a central body with a regulated objective of delivering efficient prices. This role can be referred to as *central pricing*, drawing an analogy with central planning. The monopolistic nature of access provision makes central pricing unavoidable – just as central planning of transmission is unavoidable. In a fully de-centralised and competitive market, prices would be set competitively and central pricing would not be needed.

Central pricing suffers from the same weaknesses as central planning, based around information, expertise and agency. These weaknesses are manifested primarily in relation to the baseline forecasts on which the pricing model relies.[137]

Information on the future state of the market is not available. Forecasting expertise is required to project forward from the existing state. The pricing body may not have this expertise, but through appropriate delegation and consultation it can draw on the expertise that resides in the market.

Conclusion

Concerns which have been expressed around central *planning* under the OFA model are probably better described as central *pricing* concerns. Such concerns are legitimate and largely mirror analogous concerns around central planning. However, given the monopolistic nature of transmission networks, central pricing – like central planning – is unavoidable.

More importantly, OFA is a shift to a more decentralised planning model; while forecasts are reflected in the pricing, generators have to make a decision as to whether

137 By contrast, a large amount of information on the current state of demand, generation and transmission is freely available and the expertise required to process this data is not likely to be a significant constraint.

or not to purchase access, following this pricing signal. This, therefore, represents more decentralised planning.

Central pricing concerns add to the concerns discussed previously around the robustness and transparency of access pricing. However, when considering the efficiency of planning in the NEM, and the level of co-optimisation between generation and transmission, it is appropriate to compare *central* pricing under OFA with no access pricing *at all* under the status quo, rather than with the *unachievable* ideal of competitive access pricing.

6.3.4 Thermal Pricing Model

Flowgate Support

Section 6.2.3 described how any *negative* LRIC calculated for an individual element would be *ignored* in the calculation of the access price. This section explains that design decision.

The element LRIC will be negative when, in relation to that element, the adjusted cost is below the baseline cost. That is most likely a reflection of the fact that there is negative incremental usage on the element[138] and so flows are lower in the adjusted scenario than in the baseline scenario, allowing baseline expansions to be delayed.

If an analogous situation arose in dispatch, we would refer to the generator as providing flowgate support: relieving congestion by causing flowgate usage to be reduced. Negative element LRIC can similarly be thought of in terms of flowgate support.

Flowgate support is only provided when a generator is dispatched. The pricing model implicitly assumes this, but this does not reflect reality.[139] Firm generators are not obliged to provide flowgate support and so access pricing should not credit them as though they are. This is why negative element LRICs should be ignored.

Where generators are *genuinely* able and willing to provide flowgate support, and the TNSP will save on expansion costs as a result, there is an opportunity for the generator and TNSP to enter into a flowgate support agreement, as discussed in section 4.3.3. A generator will then receive a negotiated fee for providing flowgate support which is analogous to, and a substitute for, the negative LRIC that the pricing model has ignored.

138 Meaning that the incremental usage is in the opposite direction to the flow direction in the base year. More complicated explanations are possible when the baseline flow direction reverses over the study period.

139 Recall the caveat in section 6.2.2 that firm generation dispatch feasibility is only analogous to FAPS if flowgate support generators operate at their registered access level. This assumption has been introduced to make access pricing more tractable.

It should be noted that the pricing model will *not* reflect existing or anticipated flowgate support agreements. As noted earlier, it *implicitly* assumes that *all* existing firm generators, and *no* non-firm generators, provide flowgate support.

Renewal Rights

Recall from section 6.2.3 that a renewal is included in the baseline forecast if the associated existing access has a renewal right. That right would specify how long the firm access is forecast to be renewed for. A renewal right, therefore, has the effect of extending the expiry of the existing firm access, as far as the baseline forecast is concerned, for that specified renewal length.

There are two conflicting objectives in relation to renewal rights:

- to ensure that access prices are unbiased; and
- to promote competition by ensuring that the price of an access request does not depend upon who is requesting it – other things being equal.

The first objective requires that baseline forecasts are accurate and unbiased. If it is considered likely that a firm generator will renew its access, unbiased pricing is best served by including that renewal in the baseline.

It is anticipated that, when purchasing firm access, generators are likely to request an access term which matches the expected remaining life of the associated generating asset. This is because:

- this minimises risk for the generator: if it purchases a shorter term, it will then have to renew at some uncertain price;
- *sellback* provides an exit mechanism: if the generation closes earlier than expected, the generator can sellback the remaining term to the TNSP;[140] and
- where the access request prompts a lumpy expansion, and where flow growth on that element is low, the access price will incorporate that expansion cost, irrespective of the requested term; so extending the term of the access request may be relatively cheap.

If the generator has forecast the station life accurately, the station closes at the same time as the access expires and there is no need for renewal. If the station has an extended life, renewal is possible although, at that stage, given that the remaining life expectancy may be uncertain, the generator may be reluctant to make a further purchase of long-term firm access.

Renewal rights clearly conflict with the second objective, since the renewer is offered an LRDC-based price whereas other generators are offered the (generally) higher LRIC. This particularly disadvantages new entrants who, by definition, will not have renewal

140 The sellback mechanism is discussed in chapter 7.

rights. In this respect, renewal rights would create an entry barrier that could diminish competition.

In conclusion, renewal of purchased firm access can be considered relatively unlikely and allocating renewal rights – and potentially creating entry barriers – seems inappropriate and unnecessary.

On the other hand, renewal rights will be provided with transitional access, for reasons discussed in chapter 9.

Anticipated Access Requests

The defined difference between *anticipated* access and *modelled* access is that the former anticipates a *specific procurement decision*, whereas the latter simply reflects generic forecast growth of firm access. Access procurement might be anticipated because the request has already been queued or because it is known that a new power station is being planned and is expected to require firm access.

The problem with anticipated access is very similar to the problem with renewals. Specifically, if an anticipated request is included in the baseline then it must be removed from the baseline if and when the anticipated request is received. Otherwise, it will be doubled up in the adjusted scenario. This means that, like a renewal request, the particular request is charged LRDC rather than LRIC. In a sense, the anticipated request is granted a renewal right, even though it is not renewing.

If requests can be anticipated accurately, their inclusion should improve the accuracy of baseline forecasting. But it is unclear how this forecasting would be done. New generation projects are announced relatively frequently, but only a minority are built. Including queued requests as anticipated requests in the baseline might prompt access requests that are not genuine, but submitted solely for the purpose of prompting the addition of the anticipated access to the baseline; it would be impossible to know whether a submission is in good faith.

There would also be the problem of matching access requests to their anticipated access equivalent. If a submitted access request were slightly different to an anticipated request – in location, amount or term – some judgement would be needed to decide whether this were, indeed, the anticipated request.

To conclude, including anticipated requests in the baseline forecast creates some significant forecasting, pricing and governance problems and it is not clear whether it would necessarily improve baseline forecasting. For this reason, these requests are *not* included.

6.3.5 Alternative Pricing Methodologies

Overview

The OFA model uses an LRIC-based methodology, as described above. There are two alternative approaches, referred to here as long run marginal cost (LRMC) and deep connection charging (DCC), which are used in other electricity markets and in other contexts. These alternatives are described and assessed in the sections below.

LRMC

An LRMC approach is similar to LRIC, with the essential difference that expansion *is* assumed to be *continuous* rather than *lumpy*. LRMC assumes that, if an additional 233MW, say, of transmission capacity is required, *exactly* 233MW will be built. In contrast, under the LRIC model a lumpy stylised expansion – of 500MW say – would be added.

If there is no lumpiness, there should be no spare capacity: capacity is always added in the exact quantity required.[141] Therefore, on each network element, the amount of capacity that must be added exactly equals the incremental usage from the access request. The cost of this is the product of the expansion unit cost (in $/MW) multiplied by the incremental usage. The expansion is required immediately and so the cost is not discounted. The expansion unit cost will vary, depending upon the element type. Nevertheless, the calculation of LRMC is trivial compared to LRIC. The only complex calculation is to determine participation factors, used in calculating incremental usage.

Because lumpiness and existing spare capacity are ignored under LRMC pricing, the access charge at a node where there is plentiful spare capacity will be the same as the charge where there is no spare capacity, other things being equal, despite the *true* incremental cost of transmission being much higher at the latter location. New generators may, as a result, choose to locate in locations that are relatively close to the RRN, in electrical terms (and thus have low LRMC) but are already congested, rather than choosing a more efficient, uncongested location remote from the RRN (where the LRMC is high).

Recall the three pricing objectives discussed in section 6.3.1: robustness, transparency and automation. LRMC provides a very high level of transparency and automation, due to its intrinsic simplicity. However, robustness is poor. It is clear that prices will be strongly biased in many cases: substantially over-estimating expected costs when spare capacity is high; under-estimating expected costs when spare capacity is low.

Deep Connection Charging

A deep connection charging (DCC) approach levies only the *immediate* costs of transmission expansion through the access charge and takes no account of *future* costs

141 Strictly speaking, this means also assuming *continuous disinvestment* in a negative growth situation.

as a result of future expansions being advanced. A deep connection pricing model could either be based on a stylised model similar to the LRIC model, or could be based on the cost of an actual, bespoke expansion designed by an expert planner. If the latter approach is taken, the generator might legitimately request that the smallest possible expansion is made. Such a small size may be inefficient in not providing for expected future growth.

Under the DCC, the cut-off year – the last year in which expansion costs are included in the price – is quite arbitrary. It could be set at the base year, or alternatively be set one or two years after that. Whatever the cut-off, a generator will aim to request an access volume which only brings forward major expansions to just *after* the cut-off. The generator will make maximum use of existing spare capacity, but leave as little as possible for future access requests.

The DCC methodology is similar to the existing charging approach for (shallow) connections and that makes it superficially attractive. However, the analogy between deep connection and shallow connection is weaker than the terminology might suggest. In a *shallow connection*, the connection assets are sized to reflect the needs of the particular generator: connection assets are not usually shared. Whilst there is a certain amount of lumpiness (eg, in the size of the substation bays), costs will nevertheless be reasonable proportionate to need.

A "deep connection" (ie, the immediate expansion needed to service the access request) could see expansion of elements quite remote from the generator node, for which incremental usage is quite low (but sufficient to prompt expansion), and on which lumpiness means the expansion size and cost is disproportionate to the incremental usage. As a result, DCC is likely to give rise to prices that are arbitrary, disproportionate and contentious.

In the light of these arguments, DCC seems unlikely to achieve any of the three pricing objectives.

Forecast Growth and the Value of Spare Capacity

Any new access will change the amount of spare capacity on an element. Spare capacity is likely to increase if a lumpy expansion is prompted; it will decrease if no expansion is prompted.

Although spare capacity is, by definition, *currently* unused, it will generally have some value due to the possibility of it being used to provide some *future* access. With discounting, this (present) value depends upon how quickly that future use occurs which, in turn, depends upon the current amount of spare capacity and the anticipated rate of flow growth. If spare capacity is high and/or flow growth low, future use will be distant and so net present value low.

The essential difference between the LRIC and DCC is that LRIC charges (or credits) the new generator the value associated with any reduction (or increase) in spare capacity, whilst DCC does not. Put another way, DCC assigns zero value to spare capacity. This can be illustrated using two scenarios.

Suppose that a relatively small access request prompts an immediate lumpy expansion. The LRIC will generally be less than the expansion cost. In a sense, the TNSP charges the full expansion cost to the access purchaser but then credits them with the value associated with the *increase* in spare capacity. Under DCC, the full cost is charged: no value is attributed to the new spare capacity.

Suppose, alternatively, that no immediate expansion is prompted. LRIC then charges the value of the spare capacity that is now being used by the access purchaser. Under DCC, the charge is zero: again, no value is attributed to spare capacity.

Alternatively, one can think of the LRIC method as ensuring that all generators pay for the capacity that they use for access, whether that capacity is developed especially for that generator (in the case of immediate expansion) or was provided by an earlier lumpy expansion. The cost of the expansion is shared across several generators, purchasing access at different times. Under DCC, the full cost is allocated to the generator unfortunate enough to *trigger* the expansion.

As a special case, if spare capacity on an element is estimated by the LRIC model to have zero value, then LRIC and DCC give identical outcomes. Spare capacity will have zero value in a zero or negative growth scenario in which there are no baseline expansions. Put another way, DCC *implicitly* assumes zero growth. DCC prices can then be improved upon by incorporating a more realistic growth forecast.[142]

This perspective, of valuing spare capacity, can be extended to LRMC. In a way, LRMC always values spare capacity at the unit expansion cost. If there is no expansion, the depletion of spare capacity is charged for at the unit cost. If there is a lumpy expansion, LRMC charges the unit cost on the expansion lump and credits for the extra spare capacity at the unit cost.

The value of spare capacity approximates to the unit cost when growth is high.[143] So, LRMC implicitly assumes a high growth rate. LRMC may be accurate when the forecast growth rate is high but will be biased when the forecast growth rate is low.

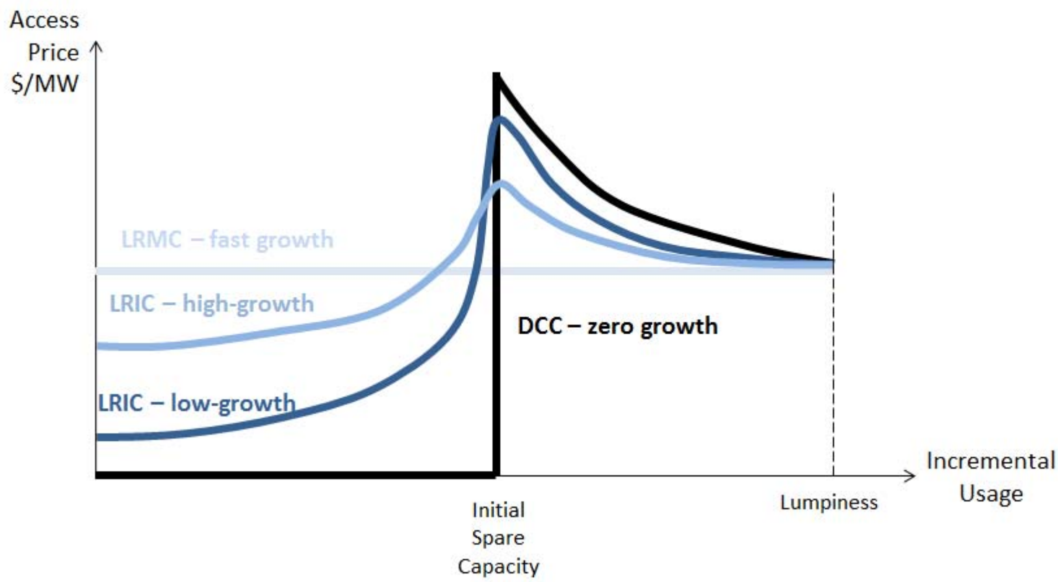
Graphical Comparison

Figure 6.3 presents the above concepts graphically. The graph plots the average access price (access price divided by volume) arising on a single element against the incremental usage on that element. The four curves show DCC, LRMC, LRIC under low-growth scenario and LRIC under a high-growth scenario.

142 In the special case, where the forecast is for zero growth, DCC pricing is unbiased. But, in that case, LRIC and DCC give the same outcome anyway, so there is no need to choose between them.

143 Specifically when the investment cycle - the time between expansions - is short compared to the discount rate: eg, where the investment cycle is 10 years and the discount rate 5%.

Figure 6.3 Access price vs incremental usage



The DCC and LRMC present the two extremes of shape that LRIC pricing can take: predicated on zero growth and very fast growth, respectively. As forecast growth increases from zero, the LRIC curve progressively flattens out, converging to the LRMC curve.

The *initial spare capacity* is the breakpoint for each curve (except LRMC, which has no breakpoint). When incremental usage hits initial spare capacity, a lumpy expansion is prompted. Further increases in incremental usage then add only slightly to cost, since the next expansion could be many years into the future. Although they are not shown, a second breakpoint occurs when the incremental usage exhausts the lumpy expansion, and then a third breakpoint and so on.

The average LRIC curves are *piecewise convex*: kinked at each expansion point and convex in between. This is true for an element LRIC and, by aggregation, true also for the aggregate LRIC – ie, the access price.

Conclusion

To conclude, the three pricing approaches considered here in a sense all belong to the LRIC family. DCC and LRMC are simply special cases of LRIC where the forecast growth is zero or fast, respectively. From that perspective, there is nothing to be gained in terms of pricing robustness by moving from LRIC to DCC or LRMC. This would mean throwing away a best-guess forecast and replacing it with a forecast that is *known* to be wrong.[144]

144 Except in the special case where forecast growth is *actually zero*, in which case the move from LRIC to DCC becomes irrelevant.

The simplicity of LRMC improves transparency and automation. DCC, on the other hand, is inferior to LRIC in that aspect as well. Because expansion costs are now borne *entirely* by the generator unfortunate enough to trigger an expansion, any DCC method is likely to be contentious and arbitrary.

6.3.6 Stability Pricing

Difficulty with LRIC for Stability Pricing

Under an LRIC approach to stability pricing, all stability flowgates that might require expansion over the study period – in either the baseline or the adjusted scenario – would need to be identified and modelled, where their expansion costs were likely to be material in the context of the thermal access price and its likely accuracy. For each stability flowgate, the capacity would need to be modelled: expressed as a function of the thermal network topology, the location and characteristics of *stability assets*[145] and the baseline demand and generation. Finally, stylised stability expansions would need to be defined and costed.

With these parameters defined, a pricing model could be established, analogous to the thermal pricing model. Spare capacity on each stability element (ie, modelled flowgate) would be calculated in each annual FAPS snapshot, and a stylised expansion added whenever a capacity shortfall arose.

There is no simple method for modelling stability flowgates that is analogous to the DC load flow used for thermal modelling. Indeed, there at least three different types of stability flowgate – ensuring steady-state, transient and voltage stabilities – which have very different characteristics and may create a requirement for at least three different stability flowgate models.

This is not an area that has been investigated thoroughly. However, the preliminary conclusion is that LRIC is not practical for stability pricing.

Pricing Objectives

The trio of pricing objectives – robustness, transparency and automation – were presented and discussed in section 6.3.1. Consideration of these objectives can guide the approach to stability pricing.

There are two relevant considerations in prioritising these objectives: the frequency of access requests; and the materiality of stability costs, compared to thermal costs.

Where requests are made frequently, automation becomes important. There may simply be not enough time to manually calculate prices for every request. It is considered possible that intra-regional access requests may be quite frequent, particularly in the context of:

---

145 Ie, network assets that enhance stability rather than provided thermal capacity.

- existing generators renewing their transitional access as it expires; and
- new generators choosing their location and firmness by making multiple indicative pricing requests.

On the other hand, inter-regional issuance is designed to combine all purchasing demand into a single, annual inter-regional auction.[146] In preparation for these auctions, access prices on DICs must be calculated for a range of amounts. However, there are relatively few DICs in the market, so this simplifies this preparation process.

Where materiality is relatively low, robustness becomes less important. For example, if stability costs are only 10% of thermal costs, then a 100% error in stability pricing is equivalent to a 10% error in thermal pricing.

The materiality of stability costs is uncertain. However, there are indications[147] that stability costs are likely to be more material inter-regional than intra-regionally.

On this basis, in the context of stability constraints automation and transparency are likely to be most important for intra-regional pricing; robustness is likely to have a highest importance for inter-regional pricing.

Lumpiness

As discussed in section 6.3.5, the three alternative pricing methods differ primarily in their treatment of – and response to – lumpiness. Indeed, Figure 6.3 demonstrated that if incremental usage equals the expansion size then all three methods give the same outcome. That was in the context of thermal pricing, but it holds also for stability pricing.

Indications are that stability expansions are typically less lumpy than thermal expansions, at least on the main transmission flow paths. Stability expansions are typically effected by the addition of new stability equipment, such as capacitors, where economies of scales are not as strong as for thermal assets and smaller sizes are commonly developed.

Furthermore, the aggregation of purchasing demands within the inter-regional auction seems likely to make inter-regional purchases – shared across multiple buyers – larger in scale than intra-regional purchases.

For both these reasons, lumpiness is likely to be less of a factor for inter-regional pricing.

146 See section 7.2.

147 Inter-regionally based on interconnector expansion studies; intra-regionally, based on a survey of TNSPs.

Conclusions

For intra-regional pricing, automation and transparency is of greatest importance. Lumpiness is significant, but the pricing errors introduced by ignoring lumpiness are less relevant due to the lower materiality of stability costs. These factors point towards adopting a simple *rule-of-thumb* approach to intra-regional stability pricing: maximising automation and transparency and not being so concerned with robustness.

For inter-regional pricing, on the other hand, robustness is of higher importance. Lumpiness may be less of a factor than for thermal expansions, so the concerns around DCC pricing in the context of thermal pricing may be lessened for stability costs. These factors point towards adopting a DCC approach to inter-regional pricing.

These are preliminary conclusions. Unlike with thermal pricing, no prototype pricing models have been developed to test the practicality and robustness of these approaches. Such quantitative analysis is needed before making any final design decision.

7 Access Issuance

7.1 Overview

Firm access is issued by TNSPs to market participants through three different issuance processes: long-term *intra*-regional issuance, the long-term *inter*-regional auction and the short-term auction. In addition, transitional access is issued prior to OFA commencement. This latter issuance process is covered in a chapter 9. In each of the other three processes, the purchaser pays an agreed charge to the TNSP who then provides the firm access certificate that is registered by AEMO.

For short-term firm access – whose term commences and expires within the next three years – there is no possibility of transmission expansion to support it. Its issuance is limited to existing spare network capacity. It costs the TNSP nothing to issue it, but it is not given away for free but rather auctioned to the highest bidder. All intra-regional and inter-regional access issuance, across all regions, is processed in a single, combined auction. This short-term auction will be held regularly, probably quarterly.

For long-term firm access – whose term commences after the next three years – a TNSP has time to develop any necessary transmission expansion and so there is no intrinsic limitation on issuance. To ensure that expansion costs are covered, purchasers must pay to the TNSP the regulated access price (in the case of intra-regional access) or at least the regulated access price (in the case of inter-regional access).

Long-term access issued to one purchaser may affect the access price offered to subsequent purchasers, meaning that the order in which concurrent access requests are processed becomes important. This issue can be addressed in two ways: through a queuing policy to establish the order; or by using an auction in which all requests are bid into an auction and processed simultaneously, with access then sold to the highest bidders. An auction is theoretically more efficient and is employed for issuing long-term inter-regional access. A long-term *intra*-regional auction is considered impractical and so a queueing approach is used instead.

Parties who hold firm access are permitted to sell it. This can be done through the auction or, in the case of long-term intra-regional access, by selling it back to the TNSP, at a regulated price.

Irrespective of its provenance, all firm access is treated identically in access settlement and so has the same intrinsic value and gives rise to the same firm access service.

7.2 Design Blueprint

7.2.1 Overview

Access Issuance and Sellback

This report refers variously to the *issuance, procurement* and *registration* of firm access and it may be helpful to clarify these terms. Essentially, they all describe the same process, albeit from different perspectives.

As described in section 2.2.4, under OFA a firm access register is established and maintained by AEMO. It records the details of all firm access being provided by TNSPs and it is used in the determination of access settlement payments and FAS obligations.

New firm access can be issued only by TNSPs. All new firm access must be *registered* by adding its details to the firm access register. A market participant *procures* new firm access through one of the issuance processes described in this chapter.

So, the usual sequence is:

- a market participant procures new firm access from a TNSP by paying, or agreeing to pay, the relevant charge;
- the TNSP issues the participant with a firm access certificate, containing the details of the new firm access; and
- the participant presents this certificate to AEMO, who then registers the access.

A sellback process is the opposite of an issuance/procurement process. Under a sellback, a firm generator or FIR holder agrees, with a TNSP, for its access to be cancelled (in whole or in part) in exchange for a payment. The participant would certify this sellback with the TNSP. The TNSP then presents this certificate to AEMO who would then deregister or amend the registered access accordingly.

Short-term and Long-term Issuance

Section 6.2.1 distinguishes between the *short-term* and the *long-term* in the context of transmission expansion, which is only possible (by definition) in the long-term. The long-term is taken to commence a fixed period ahead of real-time, which reflects the development lead-time for transmission expansion. For the purposes of this report, this fixed period is assumed to be equal to three years, but a different period may be decided on during OFA implementation.

In the *long-term*, a TNSP must expand network capacity as needed to ensure that FAPS is maintained: that FAPS capacity equals or exceeds TFGX. Firm access issuance adds to the cost of maintain FAPS and access prices reflect the cost of this. There is no intrinsic limitation on *supply* of new access and so issuance is limited only by the *demand* for new access at the access price. The long-term issuance/procurement processes are designed to reflect these fundamentals: non-zero access price and potentially unlimited issuance.

In the *short-term*, on the other hand, no expansion is possible, by definition. Since, there are no expansion costs associated with issuing new firm access, the access price, were it to be regulated to reflect costs, must always be zero. But, at zero price, there will be a high demand for new firm access. Short-term issuance must be restricted.

This done by reference to the FAPS; short-term issuance is restricted such that TFGX does not exceed FAPS capacity in the short-term. Any short-term issuance will cause TFGX to increase on the participating flowgates. Therefore, short-term issuance is only permitted where there is *spare capacity* on these flowgates: ie, where FAPS capacity is in excess of TFGX. Spare capacity might arise because of lumpy expansion or declining long-term FA levels. Where there is no spare capacity, there can be no STFA issuance.

In this way, short-term issuance is not permitted to dilute the firmness of access below the FAPS level: the level that the TNSP planned to deliver and upon which generators based their long-term procurement decisions.

In summary, the FAPS requirement (that FAPS capacity is no lower than TFGX) is maintained by the TNSP in both the short-term and long-term, but through different mechanisms:

- In the long-term, target flowgate capacity is determined by *existing* firm access and a TNSP must expand FAPS capacity to match the TFGX.
- In the short-term, FAPS capacity is determined by the *existing* flowgate capacity and a TNSP must *restrict* FA issuance to within the FAPS capacity.

Along with the short-term restriction, there is a corresponding *obligation* on TNSPs: they *must* offer short-term firm access when there is spare FAPS capacity to support issuance. Spare capacity may not be withheld in the short-term.

It should be emphasised that it is only the *issuance* processes that differ between the short- and long-term, *not* the firm access service itself. The firm access registry will not record whether firm access was issued through a short-term or long-term process. Access settlement treats all firm access equally; its provenance is irrelevant.

For convenience, this report will often refer to “short-term firm access” and “long-term firm access”, but these terms really mean “firm access purchased through a short-term issuance mechanism” and “firm access purchased through a long-term issuance mechanism”, respectively.

Intra-regional and Inter-regional Processes

As described in earlier chapters, there are some important differences between intra-regional and inter-regional firm access and the issuance/procurement processes are designed to reflect these. The most significant difference is that inter-regional firm access on a particular DIC may be provided to (through a FIR holding), and potentially demanded by, *any* market participant, whereas intra-regional access at a particular location is only provided to the *particular* generator located there.

Issuance Processes

In the light of these various distinctions, the following issuance processes are included in the OFA design:

1. transitional;
2. long-term intra-regional;
3. long-term inter-regional; and
4. short-term.

Transitional issuance involves the allocation and auctioning of transitional access and is described in chapter 9. The remaining processes are described in turn below. Each of these processes supports both issuance and sellback.

7.2.2 Long-term Intra-regional Issuance

Overview

The long-term intra-regional issuance process consists of the following steps:

1. A generator submits a *firm access request*, specifying the service parameters[148] of the new firm access it wishes to procure.
2. The TNSP calculates an access price for this request, using the pricing model that is described in chapter 6. It then must *offer* to sell the firm access to the generator at that price.
3. If the generator accepts the offer, the request is *completed* through the generator agreeing to pay the access price and the TNSP issuing the firm access certificate to the generator, as discussed above.
4. Alternatively, if the offer is declined, the request *lapses* and no further action is taken.

There are number of associated design requirements:

1. assuring payment by the generator to the TNSP;
2. processing of sellbacks;

148 These are referred to in section 2.2.4 and constitute the access term, amount and location.

3. updating the pricing baseline to reflect completed purchases;

4. sequencing the processing of access requests to ensure that the pricing baseline is always well-defined; and

5. establishing a queueing policy that defines the sequence order.

These requirements are discussed in turn below.

Access Payment

As part of the request completion, the generator must commit to pay the access price. This commitment takes place through two stages.

First, an access *payment profile* is specified. This payment profile will consist of annual (or possibly monthly) payment *instalments* that:

- are paid during the term of the firm access;
- equal, in net present value terms, the access price; and
- have a profile approved by the AER.[149]

Second, the generator enters into a payment deed with the TNSP that:

- requires the generator to make the payments specified in the payment profile;
- requires the generator to establish appropriate credit guarantees to ensure that the TNSP is not adversely affected if the generator defaults on its payments: eg, if it becomes insolvent during the firm access term; and
- specifies how and when the access will terminate in the event of a payment default.

Possible prudential and termination principles are discussed further in section 7.3.3.

As noted in section 6.2.3, the access price will, in some cases, reflect the costs of more than one TNSP. In this case, the generator still pays the full access price to the local TNSP who, in turn, would pass on the appropriate portion of the access payments to the remote TNSP. The payment profile for doing this could match the generator's payment profile. These payments could be covered by a payment deed, or similar agreement, between the two TNSPs.

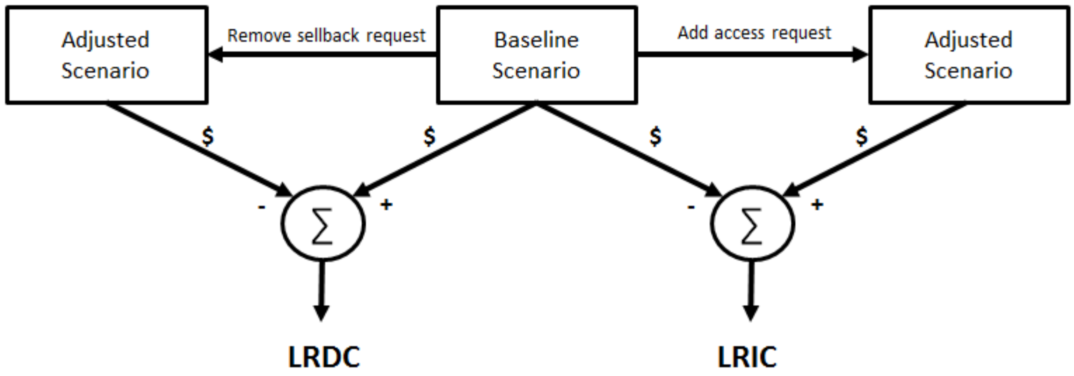
Sellbacks

Purchased firm access would come with a *sellback right*[150] which gives the holder the right to sell back all or part of the firm access to the TNSP at any time. The price for the

149 Discussed further in section 7.3.3.

sellback would be the current LRDC: this reflects the *saving* to the TNSP associated with the reduced FAPS targets resulting from the sellback. This is illustrated in Figure 7.1 below.

Figure 7.1 Calculation of LRIC and LRDC



The use of LRDC for sellbacks means that if a generator purchases some FA (at LRIC) and then immediately sells it back[151] (at LRDC), the access price and sellback price are identical[152] and the TNSP and generator are left financially unaffected.[153] The zero net payment correctly reflects the zero costs imposed on a TNSP from the two transactions.

A sellback right provides the opportunity for a generator to return its firm access to the TNSP when the avoidable cost to the TNSP of continuing to provide the access exceeds the value of that access to the generator. This promotes efficiency: it means that a TNSP will not be required to undertake expansions designed to maintain access that is no longer valued.

Since a generator, by definition, values access when it purchases it, the change of heart implied by a subsequent sellback may have two causes: either the access price has increased[154] or the access value has decreased. An access price increase could arise as a result of a forecast error in the baseline: for example, an unexpected fall in local demand due to a smelter closure creates a flowgate shortfall and prompts an expensive expansion which, when incorporated into the pricing model, causes a sharp increase in the access price. On the other hand, a value decrease might arise because a generator decides to close a power station earlier than it had expected.

Because of the profiling of the access payments, a generator that sells back its FA will have some access payments outstanding. It will still be liable for these. It will also be due a payment from the TNSP for the sellback, based on LRDC. These two payments

150 Allocated transitional access would not have a sellback right, as discussed in section 9.3.3.

151 Hypothetically. In practice, this would not be permitted, as discussed in section 7.3.9.

152 See section E.1.1

153 This result is demonstrated mathematically in appendix E.1.

154 To be clear, this means the current access price. This does not affect the generator's payments to the TNSP, which are fixed at the time of access procurement.

would be netted off (in present value terms), leaving a net amount owing to, or owed by, the TNSP. Since the generator no longer holds the firm access, it is appropriate that the net amount is settled *immediately*, rather than paid through future instalments.

Some restrictions apply to sellbacks. These are discussed in section 7.3.9.

Managing the Pricing Baseline

The role and contents of the baseline forecast are described in section 6.2.2. Importantly, in the context of access issuance, the baseline includes *existing* firm access together with forecast renewals.

When a request completes, the new firm access now exists and must be included in the baseline. This change to the baseline will affect the baseline scenario on those network elements on which the new firm access participates: ie, where it has some non-zero incremental usage. That will, in turn, affect the access price on any access request which participates on some of those same network elements.

Consider two access requests: request A and request B. If there are some network elements in which both access requests participate, then the inclusion of one in the baseline will affect the price of the other. Put another way, the price of the two requests will depend upon the order in which they are processed. If A is processed first, B will not appear in A’s baseline, but A will appear in B’s baseline, and *vice versa.*

Two access requests that affect each other’s prices in this way are said to *compete*. Some materiality threshold might be applied: eg, requests are only regarded as competing when their processing order affects their prices by more than 1%, say. The processing sequence of competing requests must be clearly defined so that pricing is well-defined and transparent. This is discussed in the next section.

Some forecast renewals are also included in the baseline.[155] So, for example, existing firm access whose term expires in 2020 may have a forecast renewal of an extension until 2025 (with the same volume and location), all of which would be in the baseline. The generator may decide not to renew, or may renew for a different term or amount. At the point that the decision is made, the renewal must be taken *out* of the baseline and replaced with the new access: or with nothing if there is no renewal. Again, this will affect the price of competing requests and the sequence of processing renewals and conventional requests needs to be well-defined.

Finally, a sellback will lead to some existing access being cancelled. Again, the baseline must be changed to reflect this, since this will affect the price of competing requests.

155 As discussed in section 6.2.2.

Sequential Processing

The sequence for processing access requests is predicated on a *queuing policy*, which is discussed in the next section. The sequence may contain requests for new access, renewals and sellbacks. These are all processed similarly.

The request at the "front of the queue" – the first in the sequence - is processed first. Until that request has been processed – ie, it either completes or lapses – it is not possible to process any competing requests, since their price depends upon the outcome of the front request. On the other hand, non-competing requests can be processed concurrently. Therefore, processing of the non-competing request that is nearest to the front of the queue can commence. If there are any requests queued that compete with neither of the two requests being processed, then the one nearest the front of the queue can also commence processing. And so on.

When the processing of a request finishes, this may free up some new non-competing requests and, again, the one nearest the front of the queue can commence processing. An example of this sequential processing is presented in appendix E.1.

Because the processing of one request holds up the processing of all queued, competing requests, the process must be time-limited. This could be done, for example, by giving the generator a limited time in which to accept and complete an offer. If this time expired, with no decision made, the request would lapse.

The sequencing restrictions only apply to the making of offers by a TNSP: ie, the calculation and offering of a *firm*, *final* price for access. A generator may, alternatively, simply be seeking *indicative* prices. These can be provided at any time, on the basis that the *final* price that is eventually offered may be different, due to baseline changes in the interim. Indeed, it may be possible for the pricing model to be made available to generators, so that they can obtain indicative prices directly without needing to approach the TNSP.

Queuing Policy

A queuing policy must be established which defines how concurrent access and sellback requests are to be ordered in the processing sequence. The simplest policies are:

- a first-in-first-out (FIFO) ("first-come-first-served") policy, based on the time at which requests were submitted to the TNSP; or
- a random "draw lots" approach, in which the ordering of a set of requests is randomised

All queuing policies are, to some extent, arbitrary, inefficient, or able to be manipulated. The extent of this problem will depend upon:

- how *long* the queue is: ie, how many competing requests are submitted around the same time; and
- the *materiality* of sequence order impacting on access prices; if this is low, then sequence order has less importance.

In the OFA design, access requests would be processed in the order in which they would be received. Ie, a first-in-first-out (FIFO) approach is applied. However, there are potentially a number of issues with this approach. Some queuing objectives and principles are discussed in section 7.3.2.

Queueing can be avoided if multiple purchase requests can be processed simultaneously. This is typically done through an auction process. Auctions are used in the other issuance processes. It is not considered feasible to use auctions for long-term intra-regional issuance, for reasons discussed in section 7.3.1.

Shape and Term Restrictions

The *shape* of a firm access request describes how its volume varies from year to year. The simplest shape is a *strip*, in which the volume is the same for each year in the access term: from commencement to expiry. In principle, any shape can be requested, and can be priced: the adjusted scenario is always the baseline scenario plus the requested shape.

In practice, there are some reasons to be concerned about the robustness of pricing of non-strip shapes. As discussed in section 6.2.4, spare capacity on a network element in the baseline scenario will have a saw-tooth shape: increasing sharply when an expansion occurs and then decreasing gradually as forecast growth uses up this new capacity. If a request had a similar saw-tooth shape, it would make use of this forecast spare capacity and would not prompt any new or advanced expansions. It would therefore have zero LRIC on that particular element.

Every element will have a different saw-tooth of spare capacity, so it would not be possible to shape and access request to avoid LRIC *entirely*. However, it may well be that shapes could be cleverly manufactured by generators to obtain a substantial reduction in the access price. This reduction is entirely arbitrary: it is unlikely to reflect genuinely low costs for the TNSP.[156] To avoid this shaping problem, only *strips* are permitted to be requested and issued.

Conceptually, the minimum term for an access request is one year. This is necessary to ensure that FAPS, which makes use of annual FAPS snapshots, is well defined. It is proposed to set the minimum term to be longer than this: possible 3 to 5 years. The minimum term is designed to achieve:

- pricing robustness; and

---

156 A monte carlo pricing approach might avoid this problem.

- a reduction in the number of access requests.

There is no maximum term.

The commencement date for access must be beyond the transmission expansion lead time: ie, three or more years ahead of the current date.

FA of shorter terms and earlier commencement dates can be procured through the short-term issuance process.

7.2.3 Long-term Inter-regional Auction

Overview

There are two differences between inter-regional and intra-regional access that are significant for the design of the issuance process:

- there are just 10 DICs on which inter-regional firm access can be procured, but there are hundreds of nodes from which intra-regional firm access might be procured; and

- there is, generally, just one generator who has access at a particular node, but there can be many market participants holding or seeking FIRs on a DIC.

These two factors mean that an auction process – rather than a queuing process – for reconciling competing requests for inter-regional access is both desirable and tractable. This is discussed further in section 7.3.1.

The design of the inter-regional auction is based on the following principles:

- bidders can submit separate annual bids for FIRs, but cleared amounts must, in aggregate have a strip shape;

- a reserve price is included which ensures that auction revenue equals or exceeds the access price of the inter-regional access that is issued;

- the sequencing of inter-regional auctions (on different DICs) and intra-regional access request processing is clearly defined;

- FIR holders can offer to sell some of their holdings through the auction, for purchase by other participants; but TNSPs are not permitted to re-purchase FIRs; and

- the auction is cleared on a platform that is developed and operated by AEMO but then settled directly between auction participants and TNSPs

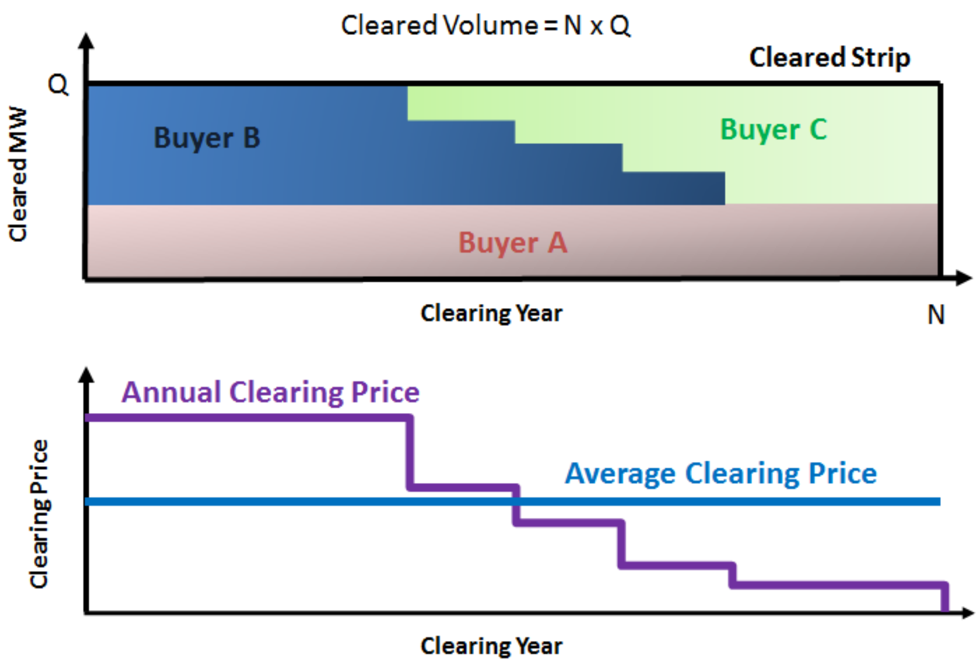
These principles are discussed in turn below. Some illustrative auction designs which would conform to these principles are discussed in appendix E.2.

Shape Restrictions

The inter-regional access that is issued through the auction must be a strip in shape, in consideration of the same concerns about pricing of shapes that arose in relation to intra-regional issuance. The pricing, and hence the restriction, applies only to the aggregate firm access that is issued in an auction, not the issuance to individual bidders.

Individual bidders need not, and should not, be subject to the strip restriction, since it is the aggregate firm access issued that impacts the TNSPs costs. It should be possible to allow each bidder to submit different bids (in price and/or quantity) for each future year. The auction design will need to ensure that the cleared amount respects both the shaped bids of individual bidders and the strip restriction for the aggregate amount. This outcome is illustrated in Figure 7.2.

Figure 7.2 Strip clearance and annual bidding



Reserve Price

The auction will include a *reserve price*[157] to ensure that auction revenue equals or exceeds the access price of the issued inter-regional firm access. Inter-regional access pricing is discussed in chapter 6. For a particular DIC, the access price would be calculated for strips of different volumes and different terms in order to establish the reserve price as a function of volume and term.

157 Which specifies the *lowest* price at which an auction is permitted to clear.

For simplicity, it will be assumed that the auction establishes a uniform clearing price for all bids cleared in each year.[158] Therefore, the auction revenue is simply the clearing price[159] multiplied by the clearing volume. So:

Auction volume x clearing price ≥ access price

Meaning that:

Average access price = access price / access volume

Therefore, the reserve price is set equal to the average access price. Typically, in an auction, the clearing price is set equal to the marginal bid price (the price of the lowest bid cleared).[160] Therefore:

marginal bid price > average access price

Since highest bids are cleared first, the average clearing price will decrease with cleared volume: it is *downward sloping*. On the other hand, the average access price will most likely *not* be *upward* sloping. The lumpiness assumptions embedded into the access pricing method mean that it is likely to have a "wavy" shape. These curves are illustrated in Figure 7.3.

Figure 7.3 Reserve price constraint

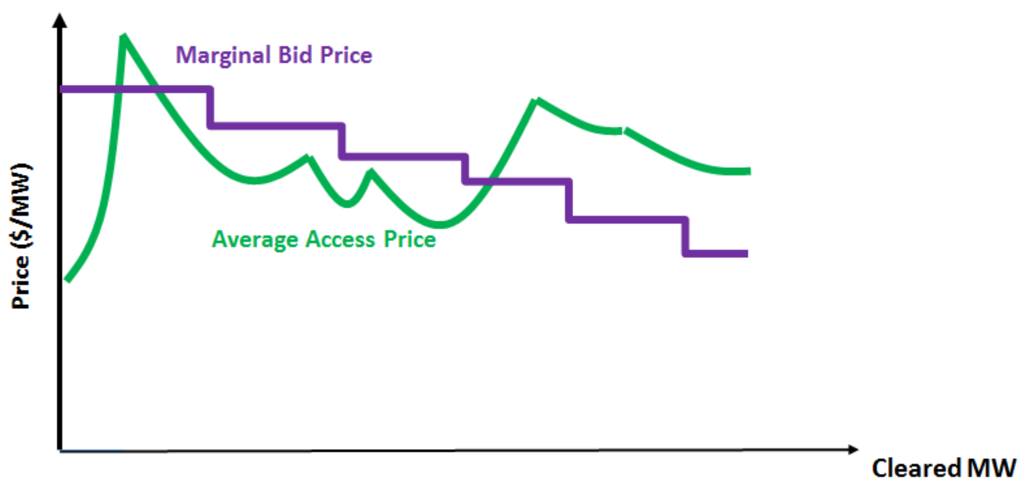


Figure 7.3 illustrates that there could be a number of ranges of cleared quantities for which the reserve price constraint is satisfied. The auction design must transparently and efficiently select a clearing point within one of these ranges. One approach would be to find the point that maximises clearing volume.

158 This is likely to be the case in the auction design, but not necessarily.

159 The arithmetical average of the annual clearing prices multiplied by the cleared volume.

160 Again, the average of the annual marginal bids is used.

On the other hand, when demand for new FIRs is low, the curve for the average clearing price may lie entirely below the curve for the average access price. In this case, the auction does not clear and no FIRs are issued.

Sequencing

As with intra-regional issuance, there is a need to sequence competing access requests[161] in order to ensure that prices are properly calculated. The major pricing competition between DICs is within a DIC *pair*: eg, Heywood eastbound and Heywood westbound. The auction would be designed to issue inter-regional access *simultaneously* on both DICs within a pair. The volume and term of access issued on the two DICs could differ. The reserve price constraint now requires that the aggregate auction revenue from inter-regional firm access sales across both DICs equals or exceeds the access price for the combined issuance.[162]

The access price for the DIC pair is likely to include some *common* costs, associated with expansions that would be required in relation to providing new inter-regional firm access on *either* DIC individually, but do not need then to be doubled up when providing access on *both* DICs: an example would be the expansion of a thermal element, located at or near the relevant regional boundary, on which there is no spare capacity in the baseline. The auction design would need to allocate these common costs between the two DICs in some way in order to establish the individual access prices and facilitate the clearing logic discussed in the previous section. To the extent that there is competition between the various DIC pairs, their auctions need to be sequenced. Since there are only five DIC pairs this is not expected to present the same degree of problem as in the intra-regional context.

There is also a need to position the inter-regional auctions within the sequence of intra-regional request processing, since intra-regional requests will commonly compete with inter-regional issuance.

Sell back

Holders of FIRs would be permitted to make offers into the auction to sell back all or some of their holdings. The auction would be designed to clear offers that were priced below the clearing price. Analogous to bids, annual offers can be submitted but issued inter-regional firm access (ie, aggregate cleared bids minus aggregate cleared offers) would *still* need to be a strip. As before, it is the *issued* inter-regional firm access that is priced and establishes the reserve price.

Although, in a sense, successful offerers would be selling FIRs to successful bidders,[163] payments would be made with the TNSPs rather than directly between auction

161 Recall that access requests are *competing* if they influence each other's prices.

162 This access price would be calculated in the access pricing model by adding the *combined* issuance to the baseline in order to establish the adjusted scenario.

163 Especially in the case where zero volume of inter-regional firm access is issued in the auction.

participants. Settlement of any sellbacks would be based on the same principles as those described for intra-regional sellback: payment would occur immediately rather than being profiled into the future; and any payments outstanding on the original FIR purchase would be netted off.

Despite this facility, FIR holders would not enjoy the same sellback rights as firm generators holding intra-regional firm access. Specifically, TNSPs would *not* be permitted to buy back inter-regional firm access. This is discussed further in section 7.3.6.

Auction Platform and Settlement

AEMO would be responsible for designing and developing an auction platform which implements the auction principles discussed above. This platform would host annual auctions. The auction would take as inputs the access price curves discussed above, as well as the participant bids.

If and when the auction clears, successful bidders will be issued with FIRs and must commit to making the appropriate payments. This could be done by requiring that all bidders, prior to the auction, enter into contingent payment deeds with the relevant TNSPs[164] which will cause an actual payment deed to come into effect when the bid is successful. The payment deed would include the same provisions as the intra-regional deed: a requirement to make annual payments to the TNSPs over the inter-regional firm access term; a requirement to provide credit guarantees; and terms around default and termination. The annual payments would, in NPV terms, equal the required auction payment and would be profiled in accordance with the same formulation as is used for intra-regional payments. Payments are directly between successful bidders and TNSPs; AEMO has no involvement in the auction settlement.

Cross-regional Settlement

Typically, two or more TNSPs would incur expansion costs associated with inter-regional access issuance on an interconnector. The aggregate of these costs would be reflected in the access price. The auction revenue should be allocated between these two TNSPs.

If the auction revenue equals the access price, it should be allocated in accordance with the breakdown of the LRIC costs by region. As discussed in section 6.2.2, thermal costs are calculated on each network element, and these can be allocated to TNSPs according to which region the element is located in. Stability costs are based on the DCC of an actual expansion plan so, again, it is straightforward to break these costs down by TNSP.

If auction revenue exceeds the access price, it is less clear how to allocate the surplus: the *auction rent*. It would probably be reasonable to allocate the auction rent according

164 The two TNSPs responsible for the DIC on which a bid is made.

to the same proportions as the access cost allocation. In any case, the proportion of the auction rent to each TNSP is then passed through to TUOS customers in that TNSP's region, as discussed in section 8.2.2.

Cross-regional settlement in relation to intra-regional access issuance is effected by the local TNSP receiving the access payments from the generator and then passing the appropriate portion on to the remote TNSP. But, for inter-regional access, it is unclear which TNSP is "local" and which "remote". Two approaches are possible:

1. The importing TNSP for a DIC receives payments from FIR purchasers[165] on that DIC and passes on the appropriate proportion to the exporting TNSP.

2. The FIR purchaser enters into payment deeds with both TNSPs and pays the appropriate portion of the auction payment to each. For simplicity, the portions could be pro rata to the overall payments due to the TNSPs from the auction.

It is not clear, at this point, which approach is preferable.

7.2.4 Short-term Auction

Overview

As discussed in section 7.2.1, access prices are zero in the short-term but access issuance is limited to existing spare capacity. A mechanism is required to reconcile potentially infinite demand (at zero price) with limited supply. An auction is both efficient and practical and is therefore the proposed approach.

The problem of auctioning short-term access is in some ways simpler, but in others ways harder, than the long-term inter-regional auction. The reserve price is zero, so the problem of calculating an access price curve – and dealing with its unusual shape - does not arise. On the other hand, there are myriad intra-regional access locations, many of which compete for the same, limited spare flowgate capacity. The auction design needs to reconcile these competing demands.

The short-term auction is designed according to the following principles:

1. issuance is limited by existing spare flowgate capacity;

2. a single, NEM-wide auction encompasses all inter-regional and intra-regional firm access, across all regions, locations and DICs;

3. there is a zero reserve price;[166]

4. firm participants can offer some or all of their firm access into the auction;

165 This reflects that, generally, the major portion of the expansion costs will arise in the importing region.

166 Ie, *negative* clearing prices are not permitted.

5. there are no intrinsic shape or structure restrictions: the auction can be designed to issue those structures that the market demands; and

6. the auction platform will be developed by AEMO, who will run, clear and settle the auction at least quarterly.[167]

These principles are discussed in turn, below.

Limited Issuance

As discussed in section 7.2.1, a TNSP is restricted in its issuance of short-term FA by the requirement that any issuance does not cause target flowgate capacity to exceed FAPS capacity: the effective flowgate capacity under FAPS conditions.

As discussed in section 5.3.2, this requirement is *similar* to requiring that simultaneous dispatch of all firm participants at their registered access level is feasible under FAPS conditions. It is not quite the same, however, because it does not allow for the possibility of non-firm flowgate support or, conversely, of firm generators *not* providing flowgate support. So the auction constraints are similar to those applying to generation dispatch, but must reflect this different treatment of flowgate support. The detailed formulation of the constraints that apply in the short-term auction is described in appendix E.3.

When firm access is issued at a node, the target flowgate capacity increases on every flowgate in which the node participates.[168] Thus, there must be some flowgate capacity available on *all* of these flowgates. This would be from a combination of:

- pre-auction spare capacity; and
- spare capacity created when a firm participant sells back some registered access through the auction.

There can be no issuance in the absence of spare capacity or sellbacks.

Single NEM-wide Auction

On inter-regional flowgates, both generators and DICs participate and so may make competing demands for spare capacity when procuring short-term access. Therefore, it is preferable to issue intra-regional and inter-regional access simultaneously within a single auction rather than issue them sequentially.[169] Since DICs participate in multiple regions, all regions must be cleared simultaneously.

167 More frequent auctions would be feasible and could be provided if there was a demand for them.

168 In the context of the short-term auction, only positive participation is relevant, since this is what determines target flowgate capacity.

169 In the long-term, access is issued sequentially rather than simultaneously because of the overwhelming complexity of calculating access prices and setting reserve prices in a simultaneous auction. In the short-term, there is no necessity to do this.

Drawing again on the dispatch analogy, this is similar to the requirement that AEMO dispatches all regions simultaneously which is needed for similar reasons.

Zero Reserve Price

As discussed in section 7.2.1, a TNSP is not permitted to *withhold* spare flowgate capacity in the short-term. The ST auction achieves this requirement by placing a *zero* reserve price on all (spare pre-auction) flowgate capacity. The absence of a reserve price means that a maximal amount of access will always be issued, such that the FAPS-related restriction means that no further issuance is possible.[170] A zero reserve price does not mean zero clearing prices, which will depend upon the level of bid demand. Where the bid demand for flowgate capacity exceeds spare flowgate capacity, a non-zero clearing price is needed to ration this demand. Plausibly, clearing prices could be quite low, if short-term firm access has low value to bidders, but they are unlikely to be zero.

Clearing prices will be driven by the scarcity value of flowgate capacity and so will vary at each location depending upon participation in the valued flowgates. However, the auction will be designed to ensure that the price of cleared bids is *never* exceeded by the relevant local clearing price: ie, bidders are never charged more than they are prepared to pay.

Sellback offers

Firm generators and FIR holders are permitted to offer to sell some or all of their holdings into the short-term auction. They would submit offer prices and the sellback would be cleared only if the relevant clearing price exceeded (or equalled) the offer price. As discussed above, the clearing price at a location is predicated on the scarcity value of the flowgates in which it participates, meaning that the clearing price will only be above zero (allowing the sellback offer, potentially, to be cleared) if there are bidders who participate in at least some of the same flowgates as the offerer. Put another way, the offer will only be able to clear if there is demand for the flowgate capacity that is freed up as a result.

The inclusion of sellback offers in the auction will enhance auction liquidity, by making it more likely that bids will clear. For example, consider a bidder at a location that participates in flowgates A and B, when there is only spare capacity on flowgate A. Without sellbacks, there is no possibility of issuing access to this bidder, since this would cause the FAPS restriction to be breached on flowgate B.

Suppose an offer is now submitted, from a different location to the bid, one that participates in flowgates B and C, say. Access can now *potentially* be issued to the bidder, since some capacity on flowgate B would be made available if the offer is

170 Although there will remain some residual spare capacity on many flowgates, where there will be no bidding generators or DICs which participate *only* in these flowgates.

cleared. Whether the auction actually *clears* depends upon the relative participation factors and bid/offer prices of the two participants.

It will be seen that the FAPS restrictions apply to *flowgates* and not (directly) to nodes. This allows for the possibility of a firm generator, in a sense, selling firm access to a buyer at a different location. A special case of this might involve an FIR holder selling firm access "to" a generator, or *vice versa*. In this way, the auction may change the relative quantities of intra-regional and inter-regional registered access. The scarce flowgate capacity is reallocated to those market participants – whether inter-regional or intra-regional - who value it most highly.

The benefit of clearing every node simultaneously within the auction, then, is that it facilitates these complex secondary trades. In a simpler form of secondary market, in which bids and offers are cleared at each node *separately*, it would not be possible for secondary trading to occur between different nodes.

No Intrinsic Restrictions on Structure

The term and shape restrictions that apply to long-term issuance relate to access pricing and FAPS obligations. In the short-term, there is no access pricing or FAPS. Therefore, such restrictions are not needed.

Although the short-term issuance is limited by FAPS capacity, which is an annual quantity, the restriction could be applied monthly, say. For example, if the short-term auction were designed to issue monthly blocks of firm access, the FAPS constraints would apply to each month *individually*. Target flowgate capacity would then vary by month, but would remained at or below the FAPS capacity across the entire year.

This flexibility opens up a new role for the short-term auction; it facilitates the *slicing* and *sculpting* of annual firm access. For example, a particular firm generator might have a greater need for firm access over the summer (say) than the winter. If it has purchased long-term, it will have purchased an annual block. The short-term auction could facilitate the sale of its unwanted winter access to another generator, who is short of winter access.

Conceptually, short-term access could be structured at a monthly, daily or even half-hourly level. In practice, access requirements are likely to be aligned with forward contracting positions and so the market will probably demand firm access structures which reflect structures in the forward market: eg, monthly and, possibly, peak/off-peak.

Auction Platform and Process

The auction platform will be designed, developed and operated by AEMO, in accordance with the principles discussed above. The inputs into the auction will be the FAPS constraints and the market bids and offers. As discussed in section 7.2.1 the FAPS constraints are essentially the same as those developed by TNSPs in their planning processes, except that they apply to the short-term rather than the long-term.

Thus, the governance of the constraint formulation would be similar to that for FAPS: TNSPs would develop the constraints in accordance with their FAPS procedures.

Unlike the long-term inter-regional auction, the short-term auction would be settled by AEMO. Successful bidders would make the appropriate payment to AEMO who, in turn, would make payments to successful offerers. Any auction surplus would be paid to TNSPs (and ultimately passed through to TUOS customers). Allocation between TNSPs would be based on flowgate tags.[171] Importantly, there will never be an auction deficit, for reasons explained in appendix E.3.[172]

Unlike for long-term sellbacks, sellers are *not* required to pay any associated outstanding payments associated with their original purchase: ie, if this were through a long-term issuance process. In fact, a seller would continue to make payments to the TNSP, in accordance with its payment deed, notwithstanding that it no longer holds the access for the short-term. This simplifies the settlement process and should not create undue credit risks, since the sellback only occurs over the short-term (long-term access not sold in the short-term auction would be unaffected).

AEMO would register all auction purchases and sellbacks in the firm access register.

The auction frequency would depend upon market demand and also on the structures being auctioned. If monthly structures were auctioned, the auction might be held quarterly or monthly.

7.3 Issues and Options

7.3.1 Auctions

Auction or Queue

Queuing is introduced to long-term intra-regional issuance because the access requests are processed sequentially and the processing order affects price outcomes. This pricing characteristic reflects the lumpiness of transmission expansion; an expansion could typically support multiple requests.[173]

The LRIC method attempts to efficiently allocate the costs of the transmission expansion between the various requests, by valuing spare capacity using the baseline forecasts. But this is inevitably imperfect.[174] A fundamental difficulty is that, in practice, the future quantity of access requests will depend upon the price, and yet the

171 See appendix E.3.

172 Unless there is negative spare capacity pre-auction, which is discussed in section 7.3.8.

173 By contrast, if LRMC pricing, which ignores lumpiness, were used the processing order would be unimportant and queuing would become unnecessary.

174 Theoretically, if the baseline forecast were perfect, every access request would have been anticipated correctly and so there would be no changes to the baseline following request completion: processing order would again become irrelevant.

price will depend upon the future quantity. It is impractical to resolve this circularity in a simple procurement mechanism.

Auctions, on the other hand, are *designed* to reconcile price and quantity in this way. In an auction, pricing is dynamic and reflects both demand and the supply constraints or costs. If capacity is scarce, the price will rise and demand will be reduced. Although auctions are still unable to reflect *future* demand,[175] they can and do incorporate all *current* demand. Instead of scarce capacity being given to the first in the queue, say, it is given to the highest bidder. In this way, an auction mechanism is always superior, in theory, to a queue.

Auction Complexity

Unfortunately, designing and operating auctions can be complex. The degree of complexity will depend upon the issuance problem that the auction is designed to solve. In the context of firm access issuance, there are two major drivers of complexity:

- the number of inter-related products; and
- access pricing.

Table 7.1 illustrates how these two drivers affect auction complexity in the three issuance contexts.

Table 7.1 Auction complexity

	Number of products	Access pricing
Long-term intra-regional	Many	Yes
Long-term inter-regional	Few	Yes
Short-term	Many	No

These two drivers are *multiplicative* in their impact on complexity. For any potential auction clearing, the access price must be calculated to verify that the auction revenue will cover it. If there are many inter-related products, the access price must be calculated as a function of the cleared volume of all of these products.[176] This makes designing an auction for long-term intra-regional issuance *far* more complex than for the other two issuance contexts. And these other two auction designs are, themselves, fairly complex.

Set against this complexity (and the associated cost) must be the benefit of using an auction compared to a queue. A theoretical benefit exists, as discussed above, but the

175 Which, by definition, is not revealed in the auction bidding.

176 The problem is also made much harder by the non-convexity of the access pricing function, discussed in more detail in appendix E.2.

actual benefit will depend upon how many concurrent access requests there are that need to be reconciled. At the extreme, if there is just *one* request, the auction will give exactly the same outcome as the queue: access will always be available at the access price: no higher (because there are no competing requests which must be resolved by setting the price at a higher bidder's bid) and no lower (as this will be below the reserve price - set to match TNSP costs).

The number and frequency of long-term intra-regional requests is uncertain. Plausibly, a new entrant will make only one substantial request: to ensure access for the life of the power station. Remaining requests will relate to existing generation. There may be many of these, primarily related to extending the term of expiring transitional access. This is a transitional problem, and not *necessarily* one that the OFA core design needs to address.

In summary, an auction is always theoretically more efficient than a queue. In practice, it is likely to be infeasible, and possibly unnecessary, to establish an auction for long-term intra-regional issuance. Auctions are proposed for all of the other issuance contexts.

Dispatch Analogy

Section 5.3.2 discussed how the FAPS requirement is similar to the requirement that a simultaneous dispatch of all firm generation and DICs is feasible under FAPS conditions. The analogy is not exact, since flowgate support generators operate differently between the two situations. But it is, nevertheless, a useful analogy for gaining an intuitive understanding of FAPS.

Since the short-term auction incorporates FAPS constraints, the dispatch analogy can be extended to this context. Recall that the post-auction issuance – the pre-auction registered access *plus* the access issued in the auction – must comply with the FAPS constraint. In the dispatch analogy, the *post-auction firm generation dispatch*[177] must be *feasible* on the *existing* transmission network.

In the auction, a generator might bid to increase its registered access, analogous to bidding to increase its dispatch level. A *high* bid price in auction is, then, analogous to a low dispatch offer price, so the two have an *inverse* relationship. A generator offering to sell into the auction is, by analogy, content to see its dispatch reduced: it will submit a low offer price into the auction, analogous to a high dispatch offer price. In fact, as discussed in the appendix E.3, the relationship is:

Auction bid/offer price = RRP – dispatch offer price

In the analogous dispatch, RRP would be *fixed* at a high level, by adding a large notional generator or scheduled demand, bidding at RRP, at the RRN. This ensures that the auction and dispatch offer prices have a fixed relationship. This is discussed formally in appendix E.3.

177 Ie, the dispatch of all generators at the registered access level that they have following the auction.

The analogy illustrates that the short-term auction problem has the same level of complexity as the dispatch problem. Dispatch is certainly not a trivial calculation, but it is currently undertaken every *five minutes* in the NEM. The short-term auction will only take place monthly or quarterly. Therefore, implementing and operating the short-term auction is expected to be quite achievable and practical.

Conclusions

Auctions are theoretically preferable to queues. But designing an auction for long-term intra-regional issuance looks impractical and possibly unnecessary. On the other hand, an auction for short-term issuance looks quite tractable: similar in design complexity to dispatch, yet undertaken relatively infrequently.

Complexity of the long-term inter-regional auction is likely to lie somewhere between these two points. Incorporating access pricing into an auction appears challenging. But, in mitigation, only two products are issued in each auction.[178] This complexity is illustrated in detail in appendix E.2 where two possible auction designs are presented.

7.3.2 Queueing Policies

Overview

The need for a queuing policy for long-term intra-regional issuance is discussed in section 7.2.2. In optional firm access, access requests will be processed in the order in which they are received, ie, a first in first out basis.

However, there are a number of potential issues that may arise from adopting this first-come-first-served approach. These issues are discussed below.

An illustrative queuing policy is then developed that could address them. This policy is presented purely for the purposes of illuminating the queueing issues.

FIFO Queue

The simplest form of queue is first-in-first-out (FIFO): requests are processed in the order that they are submitted. In this model, a place in the queue provides some *option value*. When the request reaches the head of the queue and is processed, the generator can decide exactly what access request to make[179] and can choose whether or not to accept the offer made by the TNSP. It has *options* but no *obligations*.

---

178 As discussed in section 7.2.3, access on the two DICs in a pair will be issued together within each auction.

179 It would be possible, in principle, to limit the generator's flexibility in this respect: for example, the generator might have to specify the node or amount that it is requesting when it first submits the request. However, it may be undesirable to lock the generator into these parameters when it doesn't yet know the access price. Better to provide at least some flexibility for the generator to vary its request in response to the price offered.

There is no intrinsic cost to the generator from waiting in the queue. If a place in the queue has value but not cost, demand for a place will be high and the queue could quickly become very long. There is then an incentive on generators to join the queue as early as possible, before it grows long. If the queue opens on the day of OFA commencement, say, there then might plausibly be, within microseconds, hundreds in the queue, if requests are able to be submitted electronically.

Each of the queued requests must be processed in turn. If the requests are frivolous – pure "placeholders" with no real intention of requesting access – then processing will be trivial, but could nevertheless use up a lot of time, if the generator at the head of the queue sought to delay the processing of requests further down the queue. Furthermore, it seems likely that all existing generators will submit genuine requests with a view to replacing their transitional access when it expires. Genuine requests will take some time to process, whether or not they complete.

To maintain purchasing flexibility, a generator might make multiple submissions for each genuine request. These will be scattered along the queue. If, when the first of these requests is processed, the generator finds the price unacceptable, it can decline and still have further opportunities, more "bites of the cherry", from the requests further back in the queue.

In summary, an unmanaged FIFO queue is likely to be long and impose high processing costs, due to the need to process multiple copies of each genuine request.

To manage this, some queueing costs or restrictions may need to be designed. These could be, for example:

- establishing an *annual season* meaning that only requests with a specified base year can be queued at any particular point in time; discussed in the next section;
- restricting the number of positions in a queue: for example, one position in the queue per generator;[180] and
- charging for admission to, or waiting in, a queue: eg $100k per admittance or $1k per day queueing.

Restrictions seem reasonable and fair. On the other hand, charging is likely to be rather unfair and arbitrary and creates an additional problem of who to allocate the revenue to.

Annual Season

In the absence of any specific restriction, firm access might be purchased well ahead of commencement. Recall that the long-term timeframe is considered to commence three years out, say. So, an access request with a base year of 2025, say, could be lodged no later than 2022.

180 To be clear, in this context, "generator" means access unit. A generating company with a generation portfolio would be permitted to lodge multiple requests, one for each access unit.

Suppose that an additional restriction were introduced, stating that requests with base year of 2025 could also be lodged no *earlier* than 2022: ie, they would have to be lodged in 2022, or not at all. This would give rise to an annual season: in the 2015 season, requests with a 2018 base year would be processed; in 2016 those with a 2019 base year, and so on.

Queueing and processing could be framed within each annual season by the setting of several milestone dates for each season, eg:

- the date that the queue opens;
- the date that processing commences;
- the date the queue closes; and
- the date by which processing must be completed.

The queue opening date could feasibly be in an earlier year: eg, the 2016 queue might open on 1st July 2015. It would be possible to have no queue opening date: so requests could still be submitted at the date of OFA commencement.

The processing-related dates are designed to allow sequencing between intra-regional and inter-regional access.[181] If such dates were established, the long-term inter-regional auctions could then take place during the intra-regional "off-season".

The queue closing date is needed, to ensure that processing can be completed and that TNSPs then still have the time to undertake the necessary expansions.

Since access pricing is based on baseline forecasts, which are likely to become less accurate the further out they go, access prices are likely to be correspondingly inaccurate if pricing takes place a long time in advance of the base year. An annual season means that this lead time, and the associated inaccuracy, is minimised.

An annual season does not solve the FIFO queuing problem discussed in the previous section. One might still get the situation of a large number of requests being lodged within micro-seconds of the official queue opening. The order in which these requests arrive may be quite random – reflecting the speed of transmission through the internet for example – or may reflect the skills or capacities of the various generators: effectively imposing an inefficient cost on the queuing of requests. An alternative approach, *random sequencing*, is considered in the next section to address these difficulties.

With an annual season, there can only be one genuine access request per access unit. The base year is fixed. Multiple requests might be submitted with different end years, but that would really be an attempt to circumvent the "strips only" restriction.[182] So, restricting generators to a single request per access unit could mitigate gaming of the queuing policy and issuance restrictions without preventing the lodgement of

181 As discussed in section 7.2.3.

182 See section 7.2.2.

legitimate requests. Once its one request had been processed (whether completing or lapsing), the generator might then be permitted to submit another request, so long as this was before the queue closing date.

Random Sequencing

Random sequencing is a possible alternative to a FIFO queue. This could colloquially be described as "drawing out of a hat". Out of all of the requests that are queued, one is chosen at random to be processed first, another chosen at random to be processed second, and so on.

With random sequencing, the lodgement time becomes unimportant. With an annual season, one could design a system such as:

- random sequencing of all requests lodged *before* the official queue opening; and
- FIFO sequencing of all requests lodged *after* the official opening.

The random queue would be processed first, so there would be no incentive to lodge microseconds after the official opening; just as good to lodge at any time *before* the opening.

The major weakness of random sequencing is – well, its randomness. By pure chance, a generator might find itself $1m, say, better off, or worse off, depending upon how the random order translated into access prices. Access issuance would become tantamount to a lottery.[183]

Renewal Requests

As discussed in section 6.2.3, renewal requests are access requests that relate to replacement of expiring firm access that holds *renewal rights*. These renewals are anticipated in the baseline forecast and so:

- if the renewal completes, as forecast, the baseline is unchanged; and
- if the renewal lapses, the forecast renewal must be removed from the baseline.

So when a renewal is processed, the baseline firm access forecast either reduces or remains unchanged. In contrast, for a usual access request, the baseline either increases (if it completes), or remains unchanged (if it lapses).

More often than not, an increase in baseline access will cause an increase in the price of competing access requests. Suppose that a renewal and a competing non-renewal are queued. Since the baseline can only rise, or remain unchanged, when the non-renewal is processed, the renewal generator would prefer to be processed *first*. And, since the baseline can only fall, or remain unchanged, when the renewal is processed, the

183 Although, as discussed above, FIFO may well be effectively random too, under similar circumstances.

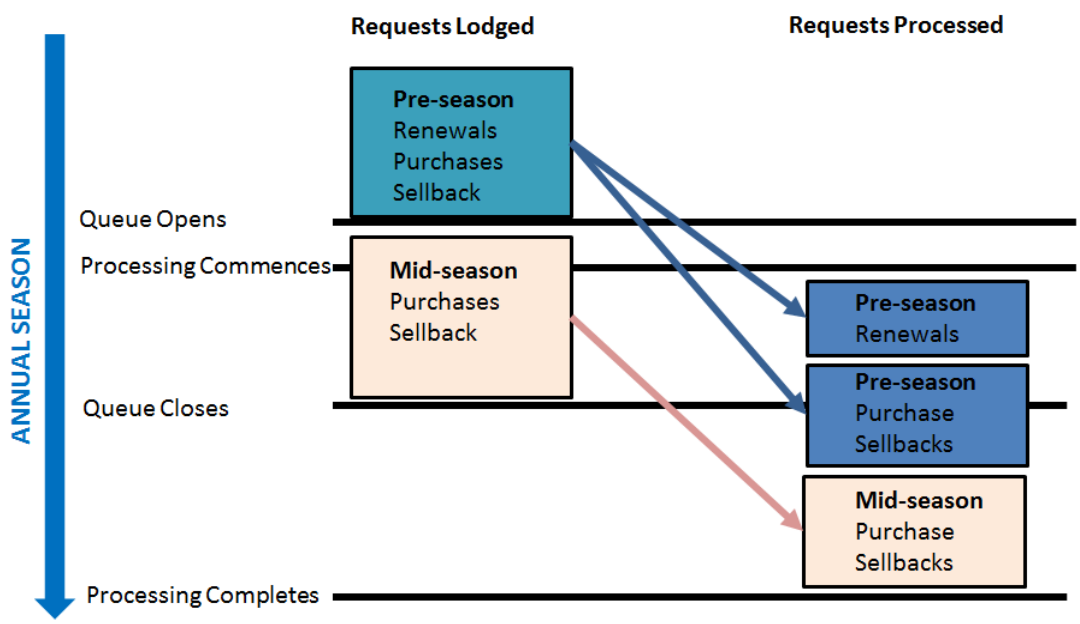
non-renewal generator would prefer to be processed *second*. So, in this case, there is no conflict: both generators would prefer a queuing order of renewal first, non-renewal second. The queueing policy could be designed accordingly.

This could be done in the context of the annual season, by setting up a queuing policy as follows:

- *Stage one*: process all *renewals* submitted before the queue opens, in randomised order;
- *Stage two*: process all *non*-renewals submitted before the queue opens, in randomised order; and
- *Stage three*: process all non-renewals submitted *after* the queue opens, in FIFO order.

In addition, to avoid the problem of the baseline being inflated by forecast renewals that do not take place, the renewal rights could be removed from any generators who have not submitted and completed renewal requests in stage one: ie, the renewals would be removed from the baseline forecast. Any requests for term extension of existing access that were submitted *after* the queue opening would therefore be treated as *non*-renewal requests.

Figure 7.4 Annual season



Slicing

Slicing is a possible approach to reducing the uncertainty associated with a randomised queue. Each queued request is sliced into multiple smaller requests: eg, a 100MW request is sliced into five by replacing the request with five 20MW requests.

The sliced requests would be processed one slice at a time in the following order:

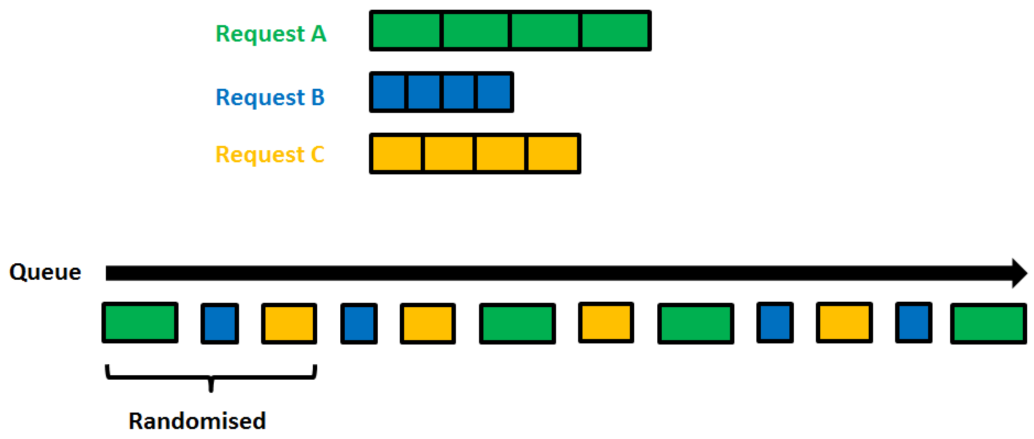
- a *first round*, in which one slice from each request is processed, in a randomised order;

- a *second round*, in which a second slice from each request is processed, with the ordering predicated on the first round order in a way which mitigated the uncertainty. For example, a generator that was randomly placed last in the first round could be placed first in the second round to offset this misfortune; and

- *further rounds* similar to the second round until all slices are processed.

This process is illustrated diagrammatically in Figure 7.5.

Figure 7.5 Request slicing



Obviously, slicing is costly in the sense that the number of renewal requests that the TNSP must process is multiplied. Set against this cost is the benefit of reduced uncertainty. By choosing an appropriate number of slices (possibly one), the costs and benefits can efficiently be traded off.

There is an obvious way for generators to subvert the slicing mechanism. A generator requiring 50MW could put in a request for 250MW, so that this is converted into 50MW slices, say. The generator then simply gets five opportunities to purchase its full requirement.

However *renewals* could be *volume-limited*, by limiting them to the registered level of the expiring firm access that they are replacing. In that case a generator renewing 100MW of TA would not be permitted to submit a renewal request in excess of 100MW request; if a renewing generator genuinely required more than 100MW, it would need to also submit a separate, non-renewal request. This restriction would prevent a generator subverting the slicing process by submitting an inflated volume, making slicing a feasible approach for processing *renewals*, in a randomised queue.

Confidentiality

A question arises about the confidentiality of access requests. A generator may be commercially disadvantaged if details of its access request are known. For example:

- if details of a request for an indicative price are published, a second, competing[184] generator might pre-empt the first by quickly lodging its own request and so be ahead in the (FIFO) queue;
- if a generator being processed knows the details of access requests behind it in the queue, this may influence its decision; and
- generators might even engage in disinformation by submitting access requests that they have no interest in completing.

In the light of this, it would seem that gaming and uncertainty is minimised by maintaining *strict confidentiality* of access request information. On the other hand, generic information about procurement queues could perhaps be published without compromising confidentiality. For example, a generator could be informed about its position in a queue so that it can be properly prepared when it is its turn to be processed.

When a request is completed, the purchased firm access is registered. This impacts the pricing baseline, the FAS requirement and access settlement. To ensure that these processes are transparent, all information in the firm access register *should* be in the public domain. Therefore, information on the access request is published if and when the request completes. The information would remain confidential if the request lapsed.

A lapsed renewal, or the cancellation of renewal rights for other reasons, affects the pricing baseline. However, it is expected that baseline forecasts will in the public domain anyway, for transparency reasons. Therefore, there is no need to explicitly publicise the fact that a renewal has lapsed. It is possible that this might be *inferred*, through analysis of changes to the published baseline.

Conclusion

There are many issues that need to be managed in a queuing policy. This section canvases some of these and discussed possible approaches and mechanisms, in the context of long-term intra-regional issuance. However, it should be emphasised that these are presented for discussion purposes only, to illustrate the queuing issues. Currently, the design decision is that all access requests would be dealt with on a FIFO basis. It is not yet clear how significant queueing issues would be if OFA were to be implemented.

184 Competing in the price sense.

7.3.3 Prudential Issues

Payment Profile

As described in section 7.2.2, when a request completes, a payment deed is drawn up and executed. The deed will specify an instalment schedule through which the access charge is progressively paid by the generator. The schedule will be based on a *payment profile*. Questions arise as to whether and how the profile should be regulated and, if so, what profile the regulations should specify.

At face value, it appears that a TNSP should be indifferent to the payment profile. Any instalment schedule will, when discounted by the TNSP's regulated WACC, have a present value equal to the access charge. If the regulated WACC reflects the TNSPs *true* WACC, the TNSP is indifferent to whether it receives the payment up front or whether it receives the payments later and has to borrow – at its WACC – to obtain funds in the meantime.[185]

On the other hand, a generator is likely to prefer a *back-ended* profile: one in which payments are made later rather than earlier. A generator will typically have a higher WACC than the TNSP, because of its full exposure to market risks. If it had to borrow to fund an upfront payment, this would be more expensive than the WACC-based discounting allows for.

This analysis does not take into account the cost of *prudentials*. The payment deed would require the generator to lodge credit guarantees so that the TNSP would not be financially impacted were the generator to default on payment: eg, because of insolvency. The cost to the generator of providing these guarantees would reflect the perceived risk of default and the amount and duration of the outstanding amounts. The cost will therefore be higher the more risky the generator's business (and so the higher the generator's cost of capital) and the more back-ended the payment profile. Theoretically, the higher prudential cost should exactly offset the lower present value (when discounted at the generator WACC) meaning that the generator, too, is indifferent to the payment profile.[186]

In any case, both the TNSP and the generator may have particular reasons in practice to seek a particular profile which this simple theory does not reflect. That would mean that, in the absence of regulation, the TNSP and generator would need to negotiate to find a mutually acceptable profile. This would take time and could delay request completion, possibly holding up those in the queue. The negotiated outcome would reflect relative bargaining power, which could distort the outcome if power is imbalanced. To address these concerns, it is appropriate to *regulate* the profile.

---

185 The design of revenue regulation could still make the profile important to the TNSP, because it impacts how much is passed through to TUOS customers. However, it is proposed to design revenue regulation to remove this dependence. This is discussed in section 8.2.2.

186 Although the provision of a forced sellback on termination, discussed below, might affect this analysis.

The AER would have responsibility for this. It could specify the payment profile that should be used. AER approval would be required if the two parties wished to use a different profile. One possible regulated profile could be to mimic the depreciation profile used on the regulated asset base in revenue regulation: eg, amortising the charge using straight line depreciation over the access term.

Termination

In the event of a payment default, it would be appropriate for the firm access to be cancelled and removed from the register; otherwise, the defaulting generator would continue to receive payments from access settlement. It may be appropriate for this to be done through a *forced* sellback process,[187] meaning that the defaulting generator receives a payment based on the current LRDC-based access price. This payment would be netted off the amount outstanding on the original access charge.

If this approach is used, the credit guarantees discussed above would only have to cover the *difference* between the outstanding payment and the current LRDC. In many situations, this could substantially reduce the size and cost of the guarantee.

7.3.4 Customisation

Overview

Firm access terms, pricing and issuance are all highly standardised and automated in the OFA design. Specifically:

- the access price is *regulated* and based on an *automated* calculation;
- firm access terms and conditions are *standardised* and are *regulated* through the FAS; and
- a TNSP is *obliged* to offer firm access at the *regulated* access price when this is requested, subject *only* to the transmission planning lead time.

The reasons for having this high-level of standardisation have been discussed at various points in this report and include:

- the need to have a single specified level of access firmness, reflecting a single FAS which, in turn, is a consequence of there being just a single, shared transmission network;
- the need for transparency and certainty in pricing and issuance;
- preventing the relative bargaining power of the negotiating parties distorting efficient prices and terms; and

187 As discussed in section 7.2.2.

- permitting a streamlined and low-cost issuance process.

On the other hand, there are potential benefits from moving away from standardisation and permitting some degree of customisation. Section 6.3.2 discussed the possibility of introducing some manual intervention into pricing in order to improve pricing robustness. Possible approaches to customising the other aspects of firm access are discussed in turn below.

Customised Terms and Conditions

The current OFA proposal does not allow customisation of registered access amounts.

However, customisation of access terms need not be in conflict with a standardised FAS. Below sets out some potential ways in which customised terms and conditions could be allowed.

For example, the access profile could be customised in a way that facilitates coordination of transmission and generation outage scheduling. A TNSP and generator might agree that registered access is set to zero over a specified outage window: eg, six weeks each year. A process could be specified through which the timing of the window is mutually agreed; alternatively, the TNSP or generator could unilaterally specify the window, with some required notice period.

The generator would then have an incentive to undertake its own maintenance outages during the access window, rather than have its generation operating but unable to obtain access. The TNSP would have more certainty that it will not incur shortfall costs during its planned outages: because of the reduced TFGX and reduced congestion. The generator also benefits by having more certainty that it gets its full access when it needs it: ie, when its plant is fully available.

The customised term need not be negotiated during the access procurement process; this might be done subsequently. A discount on the access charge could be negotiated, to reflect the benefit to the TNSP.

There are a few practical difficulties with this customisation approach which would need to be resolved:

- describing the customisation in the firm access register, on which AEMO bases access settlement;

- deciding who should pay for the cost of a more complex access settlement process;

- providing transparency to other affected generators: eg, they may want to know when the agreed outage window is;

- determining the regulatory treatment of the TNSP payment to the G: whether the cost should be borne by the TNSP or its customers; and

- reflecting the agreement in the incentive scheme design: deciding how the AER takes it into account in setting the annual benchmark.

These difficulties may be manageable, although the cost of doing so may outweigh the benefits of customisation. However, there is a more fundamental issue around the *externalities* that the customisation creates: ie, the impact that the agreement has on third parties, particularly firm participants using the same flowgates as the customising generator. If outages are to be rescheduled - or target access to be profiled - this will affect the level of shortfalls that these other participants might suffer. Perhaps these participants should then, legitimately, be involved in the negotiation between TNSP and generator.

These issues arise in the context of the particular customisation proposed. They are likely to arise generally in any customisation that might be put in place. Fundamentally, the TNSP and generator are involved in discussions that affect access settlement and therefore affect third parties.

In the light of this, some possible approaches are:

- for the AER to specify the permitted scope of customisations and, possibly, to approve individual customisations: eg, if externalities are not material; and

- for customisations to be negotiated between the TNSP and all affected generators collectively, rather than just a single generator.

Delayed Commencement of Access

The long-term issuance processes give the TNSP a specified planning lead time (nominally three years) to design, develop and commission any expansions required to provide the new access in the base year. In some cases, this may not be feasible. To mitigate this, the *planning lead time* could be increased: to 5 years, say. However, this then places risks on firm participants, who may not be able to obtain timely firm access suited to their business needs.

Alternatively, the TNSP obligation could be relaxed when the need for this was demonstrated. Relaxation could occur in a number of ways:

- the TNSP dictates a later commencement date for the access: this is then reflected in the access price and the access registration;

- the firm access commences as requested, but operates at a *lower level* for the first few years: again, this would be reflected in the price and registration;

- the firm access is provided as requested, but the TNSP's FAPS obligations are relaxed: in fact, this may not need to be made explicit, if a TNSP is unable to undertake the expansion in time it is not necessarily in breach of FAPS;[188] and
- the incentive scheme is relaxed: the annual benchmark could be increased to allow for the higher expected shortfall costs; or the shortfalls on the problematic flowgate could be excluded from the scheme.

It will be seen that, as with the other customisations, some of these approaches may affect third parties: eg, generators who participate in the flowgate and bear higher shortfall costs as a result of the delayed expansion.

To avoid the TNSP delaying access commencement simply to reducing the burdens on it or to profit from the delay in incurring cost, the AER might be involved in approving any delays. Alternatively, a regulated incentive scheme could be established in which TNSP would incur financial penalties if it delayed access and so would do this only if it were unavoidable.

7.3.5 Special Procurement Situations

Embedded Generation

A generator that connects to a distribution network rather than a transmission network is referred to as an *embedded* generator. Most embedded generators are non-scheduled but a few are scheduled (or semi-scheduled) and so may wish to purchase firm access.

A scheduled embedded generator is dispatched and so AEMO must incorporate their output into constraint equations where they can affect congestion. Clearly, AEMO must *somehow* model the distribution network when estimating participation factors for embedded generators, but the OFA design need not be concerned with how this is done. As with transmission-connected generators, OFA simply takes the AEMO analysis at face value, extracting the participation factors from the constraint equations and applying them to access settlement. In this sense, the fact that the generator is embedded is transparent to access settlement.

Distribution loss factors (DLFs) are applied to embedded generators in regional settlement. Therefore, these would also be applied to access settlement, loss-adjusting both usage and entitlements with DLFs as well as the transmission MLFs.

In transmission planning, and the short-term auction, a TNSP would similarly need to estimate the participation factor of embedded generator in order to formulate the FAPS constraints. TNSPs would need to work with DNSPs in undertaking this analysis.

188 The FAPS would probably require the TNSP only to make reasonable endeavours or best endeavours to meet the FAPS requirement.

The pricing model would also need to model the impact of the distribution network somehow. In many cases, it would be sufficiently accurate, for pricing purposes, to regard the generator as being connected to the nearest transmission node.

An embedded generator seeking long-term firm access would submit an access request in the normal way and get quoted a price based on the access pricing model. If the request completed, the firm access would be registered. The registered location would be the *local distribution node*. Access settlement, FAS and pricing obligations would flow from that.

In some cases, participation factors for an embedded generator will depend upon how the distribution network is configured. Therefore, distribution operation or planning decisions could affect a TNSP's FAS obligations. This impact is similar to cross-regional impacts at transmission level and can be managed similarly.

Grouped Procurement

Generators would be permitted to form a *group* for the purposes of procurement. The group would submit a single access request, covering the aggregate requirements of the group. The request would be priced on this basis[189] and an offer made to the group.

The group members would need to decide amongst themselves whether to accept the offer and, if so, how to divide up this cost between them. Once this was agreed, the request could be completed. There would then be multiple firm access registrations: one for each group member, in accordance with its access requirement. There would also be a separate payment deed for each group member, requiring it to pay its agreed contribution. Once the access registrations and payment deeds are established in this way, the group has no further role to play. The rights and obligations of each firm generator in the group are identical to what they would have been if the generator had made an individual access request, all other things being equal.

The value of grouped procurement is that the cost of a lumpy expansion can be allocated between group members in a way that is acceptable to all of them. If the group members, instead, submitted separate applications, the access pricing model may then allocate these costs in an unacceptable way, leading to the offers being rejected.

For example, suppose that the baseline forecast has very low growth, so that access prices are similar to DCC. Suppose also that two generators are seeking access and that a lumpy expansion can accommodate both requirements (assume also that there is no existing spare capacity). If either generator applies individually, it will be charged the full expansion costs, which may be prohibitive. If the two generators apply as a group, they can agree to pay half each, which might be acceptable to each of them.

189 The request might require access at multiple nodes. The pricing model can easily price such a request, by defining the adjusted scenario accordingly.

Super-firm Access Request

Super-firm access would be requested in an identical way to firm access. The procurement process is not concerned with generator capacity: if a request were received for 500MW, it would make no difference if the requesting generator would 1,000MW, 500MW or 200MW. The access price would be the same in each case.

7.3.6 Secondary Trading

Overview

The term *secondary trading* generally refers to an existing financial right or security being sold and transferred by the holder to another party, as opposed to a *primary* trade in which the right is originated by the provider. In the context of OFA, a secondary trade would involve a firm participant selling all or part of its registered access to another party. Firm access is specific to a particular access unit and location, and so a secondary trade would involve a change in the registered access unit and also a possible change to the location. As the term secondary trade implies, the firm access is transferred to a different generating *company*. In contrast, in a *transfer*, the access unit or location changes but the responsible generating company remains the same.

A TNSP's FAS obligations depend upon both the registered location and the registered access unit. A change in location will change the participation factors and hence the TFGXs.

Alternatively, a change in access unit, even with location unchanged, could still mean a change in target firm access (and hence a change to TFGXs), if the capacity of the original or new access unit is lower than the registered access.[190]

In relation to an FIR, there is no change in FAS obligations if the FIR on a particular DIC is secondary traded.

A secondary trade may also cause a transfer of some payment or prudential obligations to the new generating company.

For these reasons, it is proposed that all secondary trades and transfers are effected through one of the existing issuance processes or (for transfers) through a TNSP-run transfer process. It is not permitted for generators to trade *bilaterally*, without TNSP involvement or permission. The following sections describe secondary trading in the context of these various processes. However, bilateral trades at the same node can be undertake outside the TNSP-run transfer process, provided the appropriate notification occurs.

190 Recall that TFGX is based on target firm access which, in turn, is set at the lower of the registered access and the access unit capacity

Long-term Intra-regional

The long-term inter-regional issuance process allows firm access to be either *purchased* from a TNSP or sold back to a TNSP. A secondary trade or transfer can be effected by *combining* a sellback with a purchase. For example, if generator X wishes to sell some access to generator Y, this can be done through X selling back to the TNSP and Y then purchasing from the TNSP. The use of LRDC for sellbacks means that the difference between the purchase price and the sellback price reflects the cost increase to the TNSP resulting from the trade. For simplicity, it will be assumed here that the two prices are equal and so the net payment to the TNSP is zero.

It is critical that the two legs of the transaction are processed *consecutively*. If a third generator, Z, had its purchased requested processed in *between* the X and Y requests, then Z's purchase might use up the spare capacity freed up by X's sellback, leaving nothing for Y, whose access price would therefore be higher. To ensure consecutive processing, the two legs could be submitted as a single, *combined request* and queued accordingly.

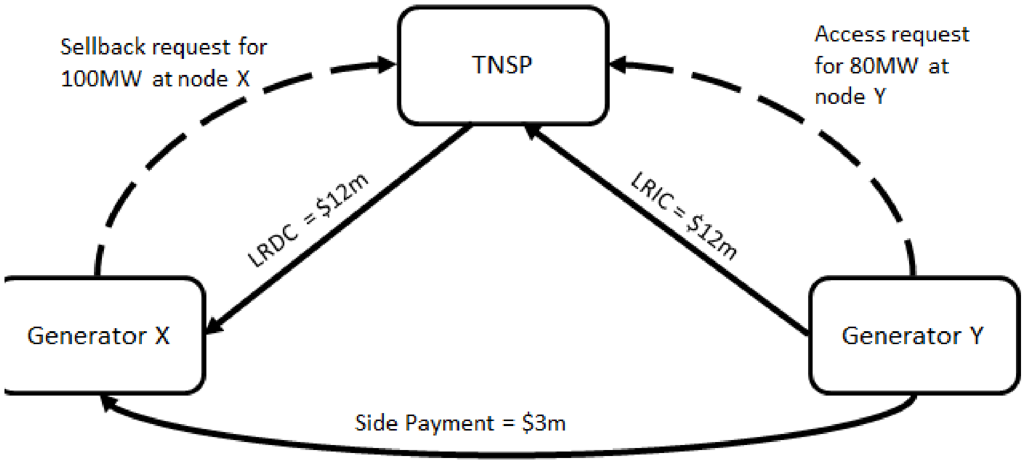
The settlement of the two legs would take place in the same way as two equivalent, but separate, requests. X would receive a sellback payment from the TNSP, but would need to make good any outstanding payments on its original purchase. Y would enter into a payment deed to pay the calculated access price.

Typically, the price agreed between the two generators would not be the same as the access price calculated by the TNSP. To correct for this discrepancy, the two generators would bilaterally agree on a *side-payment*. For example, X might have agreed to sell to Y for $15m, but the sellback and purchase prices calculated by the TNSP might be just $12m. In this case, Y would agree to make an extra side-payment to X of $3m, to reconcile the difference.

If the trade is between two different nodes, the unit access price is likely to differ. For example, the sellback price for 100MW at node X might be lower than the subsequent purchase price for 100MW at node Y. In this case, either B has to pay a higher charge, or the purchase quantity must be lower than the sellback quantity: reduced to 80MW, say. This adjustment to quantity to account for the unit price difference between nodes is informally referred to as the *exchange rate*.

An example of a secondary trade using this process is presented in Figure 7.6.

Figure 7.6 Long-term intra-regional secondary trade



Long-term Inter-regional

As described in section 7.2.3, FIR holders would be able to submit offers to sell some or all of their FIR holding into the long-term inter-regional auction. If this offer were cleared, there would typically be a corresponding clearance of a FIR bid on the same DIC.[191] In a sense, a secondary trade of the FIR has occurred, although settlement is through the TNSP, who is in effect acting as a clearing *exchange*.

Such a dual clearance will always occur where bid and offer volumes can be matched and the bid price exceeds the offer price.[192] Therefore, the auction fully supports secondary trading of this sort. Note that the auction reserve price is irrelevant in this context. If the buy and sell volumes are matched there is no new firm access issued by the TNSP and so no expansion cost. There is also zero auction revenue (net), since the payment from the buyer will exactly match the payment to the seller. In general, there could be some access issuance together with a secondary trade: eg, 200MW of FIR purchased and 80MW of FIR sold, meaning that 120MW of access is issued, which the (net) auction revenue must cover the cost of.

There is a question as to whether a *net* sellback to the TNSP should be permitted and facilitated through the auction: ie, where the quantity of cleared offers exceeds the quantity of cleared bids on a DIC. Feasibly, a constraint analogous to the reserve price constraint could be established, such that the TNSP pays *into* the auction an amount no *higher* than the LRDC of the inter-regional firm access sold back at the auction. A benefit of this facility would be that the capacity freed up on inter-regional flowgates as a result could subsequently be used for issuing new intra-regional access.

191 Conceivably, the bid could be cleared on the oppositely-directed DIC.

192 The actual clearing price for such a trade will depend upon the auction design. It will necessarily be somewhere between the bid and offer prices.

A difficulty that would arise is to how to calculate LRDC for inter-regional firm access, particularly in relation to the stability cost. Recall that the LRIC is based on DCC in this context.[193] Analogously, the LRDC should be based on the avoided DCC: ie, the amount of expansion costs avoided in the base year. This would generally be zero, since baseline expansions are typically few and far between. As a result, the LRDC might typically be low, making sellback unattractive.

In summary, secondary trading of FIRs is fully facilitated through the LT inter-regional auction. Sellback of FIRs to the TNSP through the auction is technically possible and has some potential benefits. However, it is not proposed at this time.

Short-term

Similar to the long-term inter-regional auction, firm participants are permitted to offer some or all of their registered access into the short-term auction. If an offer is cleared, there will be a corresponding cleared bid which may be at the same node or a different node. The offer and bid correspond only in the sense that one may not have occurred without the other. It would generally not be practical or meaningful to associated bids and offers in this way.

This secondary trading facility is simple and flexible. It seems probable that firm participants will primarily use secondary trading to fine-tune their access positions, and their precise access requirements are only likely to become clear in the short-term. For these reasons it is anticipated that the majority of secondary trading would take place in the short term, using the short-term auction.[194]

Transfers

Recall that firm access is registered separately for each access unit, which typically means a dispatchable unit. As discussed in section 4.2.5, this is done for practical reasons, so that values for access settlement variables such as participation factors and generation output are easily obtained. However, it does not really reflect the needs of portfolio generators, who will only be concerned with the overall access position across their portfolio.

For a portfolio generator, the particular allocation of its firm access across its access units only becomes important when target firm access become limited by the access unit *capacity*.[195] For example, consider a part-firm generator who has four access units of 500MW capacity, each with 300MW of registered access. If one unit closes, its capacity falls to zero and so its target firm access also falls to zero. The generator's *overall* target access level falls from 1200MW to 900MW as a result. To regain its

---

193 See section 6.2.5.

194 To draw an analogy, the futures market in regional electricity hedges primarily trades over the short-term. Longer-term trading is likely to occur in bilateral or over-the-counter markets.

195 The allocation of non-firm entitlements will be affected at other times, but the impact of this is unlikely to be material.

original position, the generator would need to transfer the registered access of the closed units to the three remaining units.

This transfer could be accomplished using the long-term intra-regional issuance process, as described above. Indeed, the generator would have to use this process where the transfer was between access units at different locations: in order that the appropriate exchange rate, or net payment to the TNSP, can be calculated. However, when the transfer is between units at the same location – as in the example above – this process is cumbersome and will have impacts on settlement and prudentials.[196]

Therefore, a simple administrative transfer process should be available, involving just a change to the firm access register, with no payments or prudential changes. The transfer process could be used for transfers where:

- the access units that the transfer is between are at the same node: ie, have identical participation factors on all flowgates; and
- the access units are owned by the same generating company.

Because capacity (for OFA purposes) is defined as the highest level of output over the previous two years, a unit closure should not affect the level of its capacity for some time. Therefore, the transfer process need only operate infrequently. An annual process may be most appropriate.

Synthetic Firm Access

If, for some reason, two parties at the same node – or on the same DIC - wished to trade registered access but did not wish to use the regulated mechanisms discussed above, they could use a *synthetic* firm access instrument. The seller of the synthetic right, would retain its registered access but agree to pass on all of the associated access settlement payments to the buyer of the synthetic FIR. This would lead to similar overall payments as if the registered access itself had been traded. It would be a matter for the two parties to design, agree and settle the synthetic instrument. This is beyond the scope of the OFA design.

7.3.7 Externalities from inter-regional Expansion

Overview

A principle of the OFA design is to treat inter-regional access and intra-regional access similarly, to the extent this is practical. The objective is to remove any discrimination between intra-regional trading and inter-regional trading and so promote NEM competition and cohesion. One particular design outcome from applying this is that pricing of inter-regional access is similar to intra-regional pricing: identical for thermal costs, but somewhat different for stability costs.

196 Because outstanding payments must be paid immediately, in relation to the sellback.

Commonly, inter-regional access is considered to give rise to externalities: costs or benefits flowing to parties other than the TNSP or FIR holder. These *externalities* relate to:

- reliability;
- competition; and
- regional price impacts.

An issue to be considered is whether these externalities should be recognised and reflected in the design of the inter-regional pricing and issuance processes. Or, conversely, if they are not recognised, what the consequences are for efficiency: in transmission planning in particular. These issues are discussed in the sections below.

Reliability

A major benefit of interconnectors is the improvement in reliability that they can provide: when generation capacity in a region is insufficient to meet demand, spare generation capacity in remote regions can be utilised.

Whilst this benefit is significant, it is no different in nature or magnitude to the reliability benefit from *intra*-regional access. Region boundaries, by themselves, do not change the characteristics of generation and transmission reliability.

For this reason, it is proposed that the OFA design treats inter-regional reliability benefits in the same way as intra-regional benefits. Specifically:

- the savings to a TNSP, in relation to the cost of meeting reliability standards, from new firm access are *not* included in the access price;[197]
- inter-regional reliability access may be added to the baseline forecast where this is needed to maintain reliability standards.[198] It will be a matter for the baseline forecaster to decide how to allocate any reliability shortfall between intra-regional and inter-regional RA; and
- when interconnector reliability expansions are being contemplated in a RIT-T, the benefits associated with any increase in inter-regional firm access should be included in the assessment, possibly through the use of a contingent auction mechanism.[199]

These design decisions mean that differences between inter-regional and intra-regional access are minimised, reliability standards are maintained, access prices are not distorted and transmission planning is efficient.

197 See section 6.2.1.

198 See section 6.2.2.

199 As described in section 8.2.3.

Competition Benefits

Inter-regional access promotes competition by allowing a generator in one region to supply retailers and customers in other regions or, conversely, a retailer in one region to purchase its electricity from generators in other regions. If the market structure is such that competition amongst local participants is low in a particular region, inter-regional access can improve competition by allowing participants from other regions to compete.

Clearly, the purchaser of a FIR benefits from this effect, since it becomes able to profitably compete in a market in which, due to the lack of local competition, profit margins are high. However, because this benefit accrues to the FIR purchaser, this is not an externality.

A secondary benefit is the efficiency gain associated with market prices and conditions being driven towards competitive, and efficient, levels. This benefit is an externality: it accrues across the market and not just to the access purchaser. But such efficiency benefits are *second order*[200] and so will typically be modest. If there is a benefit, it is likely to be "within the noise level" of access pricing.

Furthermore, as with reliability, it is not clear that these benefits are different in kind to the analogous benefits flowing from intra-regional access which, similarly, can enhance competition by allowing a new (intra-regional) generator to compete to supply the regional market.[201]

In summary, competition benefits are likely to be immaterial in the context of access pricing and similar in character between inter-regional and intra-regional access. For these reasons, it is not proposed to incorporate them into access pricing or issuance processes.

Regional Price Impacts

An increase in inter-regional access will generally reduce inter-regional congestion and so cause the prices in the two associated regions to converge. If one region had a generally *higher* price previously, its price may fall somewhat whilst the other region's price may rise.

There is an obvious impact of a regional price change on generators in the region. The impact on retailers is less clear: if there is retail competition, the price change will be passed through in retail prices and the impact on retailers may be limited. The impact on generators is an *externality*, since it will affect *all* generators in the regions, not just

200 The price change will cause a proportional change to market quantities. The efficiency gain is in proportion to the price change multiplied by the quantity change which, in turn, is proportional to the price change squared.

201 There was a clearer distinction, in the early days of the NEM, when generator portfolios had a regional structure, but such a structure has now largely disappeared in most NEM regions.

the generator purchasing the FIR. Of course, there is an impact on FIR holders as well, since FIR payments depend upon inter-regional congestion.

There are two, *mutually contradictory*, concerns relating to these externalities. The first concern arises from the fact that the regional price impact is simply a *wealth transfer*: the generator benefit is paid for by the retailers and customers in the region. There is no *overall* benefit to the market. A generator might fund an inter-regional expansion that benefits itself – and other generators in the region - but delivers no net benefit to the market as a whole; or, at least, delivers a net benefit which is lower than the expansion cost. This analysis implies that inter-regional access could be inefficiently *over*-provided under OFA; interconnector expansions would occur that were not economic, from a market-wide perspective.

The other concern arises from the fact that any inter-regional price convergence reduces the benefit to the FIR purchaser. At worst, if the congestion is removed as a consequence of the increased inter-regional access, then the FIR becomes worthless. Although the FIR holder's situation will have improved – since its inter-regional trading risks will have been reduced – other inter-regional traders, who did not fund the expansion, get the same benefit. They are able to *free ride* on the access increase. An FIR purchaser may find it difficult to compete with the free riders, since it alone pays the access charge. This analysis suggests that generators may be deterred from purchasing FIRs and so inter-regional access could be inefficiently *under*-provided under OFA; interconnector expansions might *not* take place, even when these *were* economic from a market perspective.

One problem with the analysis suggesting *over*-provision is that the putative generator buying FIRs to benefit from the regional price impact is similarly exposed to free riding risks. Other generators in the region benefit and yet avoid incurring the access cost. Perhaps if the generator is dominant in the region, it may be less concerned with free riders. But in that case, it could probably maintain a high regional price simply by withholding generation rather than go to the expense of funding an exporting interconnector.

On the other hand, a problem with the under-provision theory is that it takes a *short-term* view of congestion. But the generator buying inter-regional access is likely to have made a *long*-term assessment of the likely benefits. It would realise that congestion is reduced in the short-term, but could plausibly expect that congestion is likely to re-emerge in time as the market grows. The free riders, with a short-term view, will get short-term benefits only, at the expense of longer-term costs.

To conclude, there are conflicting conceptual arguments over whether the OFA design would lead to over- or under-development of interconnectors. These arguments are weak because they suppose that generators make decisions based on a short-term view which would be inconsistent with a rational and sustainable business strategy.

For these reasons, regional price impacts are not expected to distort the efficiency of interconnector expansions that are driven by FIR purchases.

7.3.8 Short-term FAPS capacity

Overview

Recall that issuance of short-term firm access relies on there being some existing spare capacity on the network: ie, where FAPS capacity exceeds TFGX. Three issues arise in relation to this spare capacity:

- timing its release over the sequence of monthly or quarterly short-term auctions;
- preventing TNSPs from hiding and withholding spare capacity, in order to reduce exposure to the operational incentive scheme; and
- dealing with *negative* spare capacity.

These issues are explored in turn in the sections below.

Releasing Spare Capacity

The short-term period lasts for three years. If short-term auctions are held quarterly, say, there will be twelve auctions that could potentially issue firm access for a particular quarter. For example, access for Q1 2020 could be issued in auctions taking place in Q1 2017 through to Q4 2019. If spare capacity were identified and offered in the first of these auctions, it will be sold (since there is no reserve price) and so there will be no remaining spare capacity, for that quarter, in any of the subsequent auctions. There are two concerns with this.

First, the release of spare capacity can improve secondary market liquidity by increasing the possibility that firm access can be secondary traded between different nodes. For example, consider the situation of a bid at a location that participates in flowgates X and Y and an offer that participates in flowgate X only. The associated secondary trade can only occur if there is some spare capacity on flowgate Y that is released in the auction. Therefore, auction liquidity may suffer in the subsequent auctions with no spare capacity being available.

Second, the estimated spare capacity is, of course, based on forecasts. If the spare capacity were over-forecast, but nevertheless fully released in the first auction, this would subsequently lead to negative spare capacity as the forecasts were corrected. This creates some difficulties, as discussed further below.

These concerns can both be addressed by *tapering* the release of spare capacity. If the spare capacity were known exactly three years out, one twelfth of this amount (say) could be released in each of the twelve subsequent auctions. However, since the forecast spare capacity is estimated and may vary between auctions, a more practical approach is:

- release one twelfth of the estimated spare capacity in the first auction;

- release one eleventh of the estimated remaining spare capacity in the second auction;
- release one tenth of the estimated remaining spare capacity in the third auction; and so on until
- release all remaining spare capacity in the twelfth and final auction.

There is a straightforward change to the auction design that ensures that only a fraction of the spare capacity is released at each auction. This is described in appendix E.3.

Capacity Withholding

The FAPS constraints used in the short-term auction are prepared by the TNSP. Revenue from the short-term auction is passed through to TUOS customers,[202] so a TNSP has no financial incentive to ensure that all spare capacity is properly identified and released. Indeed, even if the TNSP *did* retain some or all of the revenue, it might be more profitable to withhold *some* of it, in order to maximise auction revenue, if this had the effect of driving up clearing prices.

On the other hand, since a TNSP is exposed to shortfall costs under the incentive scheme, it has an incentive to *withhold* capacity in order to reduce firm access issuance and, consequently, TFGX.

In principle, an incentive scheme might be established under which a TNSP is rewarded for maximising the spare capacity released. But it is difficult to see how such a scheme would be designed and benchmarked, for several reasons. First, because the fundamental problem is that only the TNSP knows how much spare capacity there really is, it would be hard for the AER to set a *benchmark* that did not give rise to either windfall gains or windfall losses for the TNSP. Second, it is not clear how *strong* the incentive would need to be to appropriately offset the incentive to capacity withhold arising from the incentive scheme. Third, an incentive scheme would seem to *legitimise* the practice of withholding: it would seem inappropriate to have at the same time a statutory obligation *and* financial incentives on capacity release. So, an incentive scheme might actually lead to more withholding.

Finally, an incentive scheme is not really appropriate as there is no direct cost to the TNSP of releasing capacity. The capacity is already *there*, it is just a matter of entering the correct numbers into the ST auction. In contrast, under the FAOS incentive scheme, a TNSP incurs *real* operational costs in managing shortfalls.

This begs the question as to whether there is some *discretionary* FAPS capacity that the TNSP could possibly make available, but at some cost to the TNSP. For example, a network control scheme could be developed to manage the impact of contingencies and so allow a higher pre-contingent flow on a network element. If a TNSP were

---

202 See section 8.2.2.

incentivised appropriately, it might efficiently provide this discretionary capacity and release it in the short-term auction.

This incentive could be introduced by a change to the auction design, rather than through a separate incentive scheme. The TNSP would offer the additional capacity into the auction, with a *reserve price* to ensure that its costs were covered. The TNSP would then be entitled to retain, rather than pass through to TUOS, any auction revenue associated with the release of that capacity.

Such a design introduces several complexities: for the TNSP, in identifying, estimating and costing discretionary capacity; for AEMO in establishing the auction platform and settlement mechanism; and for the AER in ensuring that a TNSP is not double-dipping: asking to be paid for capacity that should already be provided from its existing regulated revenue. It is therefore not proposed to introduce such a design.

Even if it were introduced, this approach does not address the problem of ensuring that the *base* level of existing spare capacity (ie, available at no direct cost to the TNSP) is released. It is considered that this is best done through regulatory oversight and enforcement, rather than through a financial incentive mechanism.

Negative Spare Capacity

FAPS constraints in the ST auction require that the post-auction TFGX is no higher than the FAPS capacity: ie, that post-auction spare capacity is zero or positive. If the pre-auction spare capacity is *negative*, then the FAPS constraint will force the TNSP to *buy back* – rather than issue – FA in the auction. As a result, the TNSP will be required to pay money into the auction: ie, the auction revenue will be negative. The cost of this buyback could be quite high if firm participants are reluctant to relinquish their existing registered access.

Negative spare capacity might arise as a result of FAPS failure. Perhaps the TNSP failed to identify the impending shortfall in its planning process, or was too slow in developing the necessary transmission expansion. In that case, the AER might require that the TNSP remedy the shortfall through ST auction buybacks. The negative capacity "problem" might then be a suitable mechanism for penalising FAPS non-compliance.

On the other hand, there might be no FAPS failure: for example, the negative capacity might have arisen because of a reduction in flowgate capacity that could not have been anticipated by the TNSP: eg, an unanticipated smelter closure. In this case, it appears inappropriate to make a TNSP bear the cost of rectifying the problem, and so negative spare capacity in the auction should be avoided. This can be done by a correction to the FAPS capacity levels entered into the auction, according to the formula:

Auction Capacity = max(FAPS capacity, pre-auction TFGX)

If there is pre-auction negative spare capacity, this correction increases the spare capacity to zero (as far as the auction is concerned). No firm access will be issued,[203] so the negative capacity situation will not be made *worse* by the auction, it will simply be ignored.

Any negative spare capacity is likely to exacerbate TNSP penalties under the incentive scheme: if there are shortfalls under FAPS conditions there are likely to be shortfalls in other conditions too. A TNSP might plausibly volunteer to buy back access in the ST auction in order to reduce these penalties. If that were permitted, it would be effected by allowing negative spare capacity to be input into the auction. Revenue regulation would need to ensure that the associated costs were borne by the TNSP and not passed through to TUOS customers.

7.3.9 Sellback restrictions

Overview

As discussed in section 7.2.2, generators have the right to sellback firm access to the TNSP, at the LRDC price. This facility raises a possible prospect of generators *churning* access: repeatedly buying, selling back, buying back and so on. This will impose extra costs on the issuance process. But of more concern is that generators might in this way be able to exploit any lack of robustness in the access pricing model. If, for some reason, prices are quite volatile, a generator could submit repeated sellback requests and then complete the request when a high price is eventually produced by the pricing model. It could then submit repeated purchase requests until eventually a low price is offered, and so on.

Such behaviour is essentially *speculative*: the generator is buying, or selling back, firm access on the expectation that the access price will rise, or fall, respectively. The generator can then reverse its transaction and lock in a profit.

Speculation is not necessarily a negative force in financial markets. Speculators acting to arbitrage between efficient prices and actual prices will move prices closer to their efficient level. Plausibly, they could play a similar role in access pricing: correctly anticipating a smelter closure, for example, that the baseline forecaster had missed.

On the other hand, speculation substantially increases the risks relating to pricing robustness. Poor pricing could, as well as engendering inefficient procurement and expansion decisions, lead to TNSPs losing money to speculators. The TNSP is particularly at risk – compared to the analogous financial market participant – because it has no discretion: it is obliged to buy or sell firm access at the access prices determined by the AER in accordance with the regulated pricing methodology.

A number of restrictions on sellbacks are proposed, designed to mitigate this risk from speculation. These are described below.

203 To nodes that participate in the problematic flowgate.

Buy-sell Spread

A *buy-sell spread* will be added to access prices. For example, if a 1% spread were applied, generators selling back access would only be paid 99% of the LRDC rather than 100%. As well as discouraging churning, a spread aims to ensure that, were churning to occur nevertheless, the associated TNSP costs would be covered.

Minimum Holding Period

A *minimum holding period* will apply to long-term firm access. If this was set at one year, for example, generators would not be permitted to submit a sellback request until one year after the purchase date. Speculators relying on pricing volatility would only be able to “play” the pricing model at least once per year. This could significantly dampen speculation but would be unlikely to impact significantly on legitimate purchasers.

TA

It is proposed that sellback rights will not be provided on TA. The reasons for this are discussed in section 9.2.2.

TNSP Waivers

A situation could plausibly arise where a TNSP was planning an imminent expansion, which could be avoided if there were a sellback instead. In this situation, the TNSP could choose to waive the sellback restrictions discussed above.

8 Changes to existing processes

8.1 Overview

The introduction of OFA will necessitate changes to two existing mechanisms that regulate TNSPs: revenue regulation and the regulated investment test for transmission (RIT-T).

Revenue regulation under OFA will require that the *combined* revenue from TUOS service provisions and firm access service provisions is forecast not to exceed a revenue cap determined by the AER, based on the efficient cost of building and maintaining the shared network to provide those services in accordance with the relevant service standards. Revenue regulation must reflect the fact that access revenue is essentially fixed once firm access is agreed, meaning that any variations in cost or revenue from forecast must be borne by either TUOS users or the TNSP itself. The AER will be responsible for defining mechanisms for managing and sharing these forecasting risks.

Under the RIT-T currently, TNSPs are permitted to develop network expansions on the basis of the improvement in access that they provide to generators. This conflicts with the OFA objective of generators themselves deciding on their access level, and procuring firm access and paying access charges, accordingly. Some changes are required to the RIT-T to better align it with the OFA model.

8.2 Design Blueprint

8.2.1 Introduction

The overall approach to revenue regulation is the same under OFA as it is currently.

Under OFA, some *new* regulated obligations are placed on TNSPs, such as complying with the FAS and access issuance. However, OFA also impacts on two areas of existing TNSP regulation:

- revenue regulation; and
- RIT-T obligations.

Some design changes to these areas are proposed, to align them with OFA, whilst retaining existing elements and principles. These are covered in turn below.

8.2.2 Revenue Regulation

Overview

TNSP revenue regulation limits the revenue that a TNSP is permitted to recover from charges for regulated transmission services: specifically TUOS services. The regulation process, undertaken by the AER, aims to allow TNSPs to recover the reasonable costs associated with TUOS provision and to incentivise capital and operational efficiency. It does this by setting a revenue allowance based on the estimated costs for a TNSP, so that if a TNSP can reduce its costs below this level it will increase its profitability.

Currently, the majority of the costs incurred are associated with the development and operation of network assets required to deliver TUOS services in accordance with reliability standards (RS). Revenue is primarily recovered from TUOS charges on TUOS users: ie, transmission-connected loads and DNSPs. There are some additional costs (eg, network support agreements) and revenues (eg, from the Settlement Residue Auction) that are typically permitted to be *passed-through* to TUOS users: ie, the revenue allowance is adjusted by an amount equal to these costs and revenues.

OFA introduces several new categories of cost and revenue for TNSPs, including:

- the cost of developing and operating network assets required to maintain the FAS;
- penalties (or rewards) under the FAOS incentive scheme;
- revenue from issuance of long-term intra-regional access;
- revenue from issuance of long-term inter-regional access: this can be subdivided into the access price and the auction rent: the latter being the amount by which auction revenue exceeds the access price;
- revenue from the short-term auction; and
- inter-regional payments to or from TNSPs under access settlement and regional settlement.

The revenue regulation process needs to be changed in the light of these new payment streams.

Revenue Reset

Currently there is a TNSP *revenue reset* every five years. For each of the five years in the following *regulatory period*, the AER calculates an Aggregate Annual Revenue Requirement (AARR), being the prudent and efficient costs of delivering forecast TUOS volume over the period, given the existing asset base. The AARR covers the *carrying cost*[204] of existing and forecast new assets, together with forecast operating costs. A TUOS revenue cap is then defined so that in NPV terms, it provides an amount equal to the AARR.

Under OFA, two changes to the TNSP's commercial situation must be reflected in the revenue reset process. First, the TNSP must now deliver both TUOS and firm access

[204] The carrying cost of an asset is the WACC-based return on the depreciated asset value plus a depreciation allowance.

services over the next regulatory period and the AARR must represent the efficient cost of doing this. This will be predicated on forecast levels of TUOS *and* firm access. The firm access forecast would include existing firm access and any anticipated or modelled firm access. Only *long-term* firm access creates costs for a TNSP and this must be bought three years in advance. This makes the forecasting process somewhat easier.

Second, the TNSP now has two major sources of revenue: TUOS revenue and firm access revenue. The latter can be estimated from the payment profiles of existing and forecast firm access. Firm access revenue will cover some of the AARR and the TUOS revenue cap should be set to recover the remainder: ie, in NPV terms, the TUOS revenue cap is set so that:

TUOS revenue cap = AARR – forecast firm access revenue

Network assets may be built in response to reliability requirements, FAPS requirements or, commonly, a combination of both. It is therefore not meaningful to categorise network assets as "reliability assets" or "access assets" and so the regulatory treatment of all network assets must be the same. In particular, all new assets will be *rolled-in* to the regulatory asset base[205] at the revenue reset, irrespective of the reason for their development.

Treatment of Variances

Actual costs and TUOS volumes will almost inevitably turn out differently to the forecasts used in the revenue reset. The TUOS revenue cap is currently *not* adjusted to reflect these variances.[206]

Under OFA, variances will also arise in relation to the volume of firm access issuance. Suppose that some additional long-term firm access is issued that was not anticipated in the forecast. This has two impacts on the TNSP:

- the TNSP will receive some additional access revenue over the regulatory period, based on the access price and payment profiling of the additional access; and
- the TNSP may incur some additional costs associated with expanding flowgate capacity to accommodate the new access and maintain FAPS.

The variance is unrelated to TUOS and so the TUOS revenue cap is *not* adjusted. Therefore, any difference between the *revenue* and cost variance will impact the profitability of the TNSP.[207]

Recall that the access price is set at the estimated LRIC of the new access and that the pricing model is designed to ensure that the estimate is robust and unbiased. In this

205 This means that the actual cost of the assets is included, subject to an ex-post efficiency review.

206 Except under special circumstances specified by the AER.

207 Contrast this to the current situation where no TUOS *revenue* variance is permitted and it is just the TUOS *cost* variance that impacts on TNSP profitability.

respect, any revenue variance (reflecting the access price) should be similar to the cost variance (reflecting the access cost) and so any profit variance (being the difference between the two) should be relatively low and not cause a substantial increase in TNSP financial risk.[208] Therefore, the revenue regulation process is not designed to mitigate this particular risk.

Timing Mismatch

Even if cost and revenue exactly match (in NPV terms) over the *long term*, they will generally *not* match over the *next regulatory period*. This is due to likely differences between the timing of costs, driven by the expansion plan, and the timing of revenues, which is defined in the payment profile.[209]

The impact on the TNSP of this timing mismatch can easily be calculated. The timing of expansion costs is calculated as part of the access pricing process.[210] The payment profile of new access is recorded within the payment deed.

To cancel out the financial impact on the TNSP of this timing mismatch, the AER will – at the next revenue reset - *increase* the TUOS revenue cap by the amount of the timing impact. The timing impact is symmetrical.

Whilst the impact is now borne by TUOS customers instead, TUOS revenue will *automatically* adjust in future regulatory periods to leave customers neutral over the long term. This is explained in appendix F.1.

With this correction, the TNSP - and its customers - becomes *indifferent* to the payment profile, allowing the profile to be regulated rather than negotiated.[211]

Pass-Through

Other cost and revenue streams are introduced by OFA. The revenue regulation process can either ignore these payment streams or can arrange for them to be passed through to TUOS customers. Pass-through is effected by adjusting the TUOS revenue cap by an appropriate amount: eg, a new cost stream of $3m in a year would be passed through to TUOS by increasing the TUOS revenue cap by $3m. If the payment stream is passed-through then TUOS *customers* will bear the cost (or enjoy the benefit). If it is ignored, then it the TNSP *itself* that bears the payment impact, on its bottom line.

---

208 Of course, this needs to be tested empirically, but it is a working assumption for the process design.

209 It would not be suitable to base the payment profile on the timing of long-run incremental costs. Incremental costs are likely to be high during the access term, reflecting incremental expansion, and then become negative beyond the access term, reflecting the remaining value of expanded capacity.

210 This gives the timing of the capital expenditure, from which the timing of the associated carrying costs can be calculated, based on the WACC and the depreciation schedule.

211 As discussed in section 7.3.3.

The principle used to determine which approach to take is that costs and revenues should be passed through when they are *not directly related to TNSP activities or the costs incurred by those activities*.

Access revenue received by TNSPs can be divided into two components:

- a *cost* component: ie, the access price;[212] and
- a *rent* component: the amount, if any, by which revenue exceeds the access price.

As illustrated in Table 8.1 below, rent can arise in issuance auctions, where excess demand for firm access leads to clearing prices being higher than the access price. For the short-term auction, access prices are zero and so the entire auction revenue is rent. There is no rent arising from the long-term intra-regional issuance, where access payments are based directly on access prices. This breakdown is summarised in Table 8.1.

Table 8.1 Categorisation of Firm Access Revenue

Issuance Process	Cost component	Rent component
Long-term intra-regional	All revenue	Zero
Long-term inter-regional	Most revenue	Some revenue
Short-term	Zero	All revenue

In accordance with the principle stated above, the *rent* component is passed-through but the *cost* component is ignored (and so retained by the TNSP).

As discussed above, any differences between revenue and cost variance within a regulatory period, due to timing differences, are also passed through to TUOS.

Penalties or rewards associated with the FAS incentive scheme (or any other incentive scheme introduced under OFA) are *never* passed through, since if they were the incentive on TNSPs would be lost. Obviously, the penalties and rewards reflect TNSP activities and costs and so this allocation is consistent with the principle stated above.

Inter-regional Settlement Payments

The final payments stream to TNSPs relates to inter-regional settlement. Currently, TNSPs receive the auction revenue from the sale of SRA rights and also pay the cost of any negative IRSR arising during counter-price flows on interconnectors. These revenues and costs are passed-through to TUOS customers.

Under OFA, the inter-regional payment streams are somewhat different. They arise from:

212 Which is designed to be cost-reflective.

- payments due to non-firm inter-regional entitlements;[213]
- costs associated with flowgate support interconnectors;[214] and
- the inter-regional losses residue.[215]

As described previously, the first two categories of payment are allocated between TNSPs based on flowgate tagging. Losses residues arising on a DIC are allocated to the TNSP in the importing region.

The non-firm payments and losses residue are not related to TNSP activities or costs, and so they are passed-through to TUOS.

The cost of flowgate support interconnectors *is* related to TNSP activities, since it results from a weakness in the TNSP's network. But, as discussed in section 4.3.4, flowgate support interconnectors are analogous to network support generators: those generators with whom a TNSP has entered into a network support agreement (NSA), in order to maintain reliability. The costs of NSAs are passed through to TUOS.

Although there is no written agreement with flowgate support interconnectors, a service is nevertheless provided which is similarly a substitute for reliability expansion. The strength of this analogy suggests that the regulatory treatment of the two services should be the same. Therefore the costs of interconnector flowgate support are passed through to TUOS.

8.2.3 RIT-T

Current Arrangements

The RIT-T is a cost-benefit analysis which a TNSP must undertake prior to committing to any major network expansion. Through the RIT-T process, the TNSP must demonstrate that the proposed expansion provides the maximum benefit (or minimises the cost) of all *feasible* expansions: ie, expansions which maintain reliability standards.

One can conceptually distinguish between two types of expansion:

- a *reliability* expansion, without which reliability standards will not be maintained; and
- an *economic* expansion, which is not needed for maintaining reliability standards.

Since doing nothing (which has zero net benefit) is a feasible alternative to an economic expansion, the net benefit of the expansion must be positive. The net benefit of a

213 See section 4.2.4.

214 See section 4.2.4.

215 See section 4.2.5.

reliability expansion, on the other hand, might be negative (as reliability standards must be maintained).

FAPS expansions

Under OFA, a third type of expansion is introduced: a *FAPS expansion*, without which the FAPS requirements would not be met. As with reliability expansions, the *measured* net benefit from a FAPS expansion may be negative (as the FAPS must be maintained). FAPS expansions will commonly be distinct from reliability or economic expansions. However, some expansions may help in delivering both FAPS and reliability. These are not easily categorised.

In any case, the categories are for illustrative purposes only. The RIT-T requirement remains the same irrespective of category: the proposed expansion must deliver the highest net benefit of all feasible expansions. An expansion is now infeasible if it means that *either* reliability standards *or* FAPS are not maintained.

Generator Benefits under Current Arrangements

Under current arrangements, a TNSP is required to include in the RIT-T all material costs and benefits that accrue to the TNSP or to other NEM participants: ie, generators, retailers and other TNSPs. In particular, an expansion project may be chosen and justified under the RIT-T (at least in part) because of the benefits that it is anticipated to provide to generators: for example by relieving congestion.

This inclusion of generator benefits in the RIT-T allows TNSPs to provide a *quasi-economic standard* of access to generators: ie, the access level at which the total of transmission costs and congestion costs is minimised. The standard provided by TNSPs currently differs, nevertheless, from the *economic* standard that OFA delivers, described in section 5.3.1, for two reasons.

First, the RIT-T only *allows* this, it does not *ensure* it. Unlike with reliability expansions, a TNSP not *required* to develop economic expansions. Provision of the economic standard is optional, not mandatory.

Second, the economics of the expansion are evaluated by the TNSP, rather than by generators.

Under OFA, generators select their preferred firmness of access, opting to be non-firm, part-firm or fully-firm, through their procurement decisions. But if economic expansions continue under OFA, generators could actually be provided with a quasi-economic standard of access which is superior to their chosen level. This seems to breach the “sovereignty” of the generator decisions and potentially undermine the OFA mechanism in delivering transmission planning efficiency.

This issue is discussed below, in the context of non-firm generators and firm generators, in turn.

Non-firm Generator Benefits under OFA

It is proposed that, under OFA, benefits to *non*-firm generators relating to the relieving of congestion – and associated improvement in access firmness – should not be included in the RIT-T for reasons set out below.

If the benefits *were* included in the RIT-T, a TNSP could provide a quasi-economic standard of access to non-firm generators. This standard would be superior to non-firm access but probably inferior to fully-firm access.[216] So, it can be equated to some level of *part*-firm access: 60% firm, say.[217]

A generator that was content with this firmness level might opt to be non-firm: to free-ride, hoping that the TNSP will provide the part-firm standard anyway. Generators wishing to be firmer than this, on the other hand, would need to procure firm access. They would be in a position where a large part of their access price – that needed to bring them to the part-firm level – is effectively wasted and so they are paying a high access price for only an incremental improvement in access firmness. For example, a generator wishing an 80% firm service would pay for the full 80%, despite only obtaining a 20% improvement in firmness.

These concerns *could* be addressed by adjusting access pricing, so that generators are only charged for the incremental costs over and above the economic standard: ie, for FAPS expansions that would not otherwise be undertaken as economic expansions. That would be practically problematic, because the level of this quasi-economic standard is unclear, being provided at a TNSP's discretion. It is also inconsistent with the treatment of reliability expansions, which are effectively ignored in the access pricing method.

In summary, it is proposed that a TNSP should *not* be permitted to include in the RIT-T those benefits to non-firm generators that are associated with reduced congestion. This applies to both economic expansions and FAPS expansions. The RIT-T treatment of reliability expansions – and of the benefits arising from reliability access - is considered in a separate section, below.

Firm Generator Benefits

A similar issue arises in relation to the inclusion of access-related benefits – ie, the *reduction* in shortfall costs - to firm generators in the RIT-T.

If firm generator benefits are incorporated into the RIT-T, economic expansions may be developed that lead to generators getting a superior level of access firmness to that provided from FAPS expansions alone: ie, a firmness that is superior to the FAPS requirement. This is explained below.

216 Fully-firm access would mean absolutely no congestion under FAPS conditions. This is unlikely to be delivered under the quasi-economic standard.

217 A part-firm service is a blend of non-firm and firm access services that is provided to a part-firm generator. A 60% part-firm service would be a blend of 60% firm and 40% non-firm.

FAPS only guarantees access under FAPS conditions. Since these conditions are likely to arise, at most, for one half-hour per year, a TNSP could, hypothetically, meet FAPS by providing target flowgate capacity *only in this one half-hour*, and *zero* capacity at all at other times: a *minimal* – and almost worthless - level of access firmness. The FAPS is designed on the *expectation* that, if flowgate capacity is available under FAPS conditions, it is likely to be largely available for most of the remainder of the time. Generators, correspondingly, will procure firm access only if they have a similar expectation; they would not pay for the "minimal" service.

But this expectation could be misplaced, particularly where the FAPS expansion uses *alternative* technology: ie, *not* network expansion. For example, a TNSP could perhaps contract with some demand management, which enhances flowgate capacity when it is activated. To meet FAPS, the demand management need be activated only under FAPS conditions, such that only the minimal access firmness is delivered.[218]

Consider the RIT-T in this context. This minimal demand management option might be compared to another *anytime* demand management option, under which demand management is called anytime there is a shortfall. A third option might be network expansion.

The first option is likely to be the cheapest, but the second and third options will bring additional benefits to firm generators. If these benefits are ignored, the first option will be preferred as part of the RIT-T analysis. If they are included, the second or third option is likely to be preferred. Therefore, including generator benefits will commonly lead to more "firm" solutions being chosen and a firmer level of access being delivered as a result.

The access price is based on stylised expansions of network capacity. Therefore, a generator might legitimately expect to receive a service firmness that is consistent with the firmness of network capacity, rather than a poorer service delivered by cheaper technology. Therefore, unlike in the non-firm generator case, the firm generator is not really being *gifted* this superior service but, rather, receiving a service that is consistent with the price it paid for its firm access.

That is not to say a TNSP should be *prohibited* from using alternative technologies in FAPS expansions. In the above example, the anytime demand management option might deliver access firmness similar to network capacity and be much cheaper. It would be wrong to rule it out.[219]

This argument presents a strong case for including firm generation benefits in the RIT-T. On the other hand, there remains a fundamental concern that this would be inconsistent with the OFA principle that *generator* decisions drive transmission

218 It should be noted that the FAOS incentive scheme may, by itself, be sufficient to make this option undesirable, since the associated shortfalls could lead to high penalties on the TNSP. However, this incentive is diluted by the caps placed on penalties.

219 This is especially the case given the recent and ongoing rapid advances in demand-side technologies.

planning. Expansion decisions would remain dependent on TNSP estimates of generator benefits, if only in part.

There is also a practical concern. Estimating congestion costs is difficult for TNSPs, who do not have expertise or involvement in market behaviour and outcomes.[220] It will become more complex if these costs must then be allocated between firm and non-firm generators.

To conclude, some more analysis of these pros and cons is needed before a final decision on this issue is taken.

Reliability Access

The previous two sections consider whether generator benefits should be included the RIT-T in the context of economic and FAPS expansions. A similar issue arises in the context of reliability expansions.

As discussed in section 5.3.5 reliability expansions will in some cases provide *reliability access* to non-firm generators. It would be appropriate for a TNSP to include the value of this reliability access within the RIT-T in order to select the efficient expansion option. However, this would be inconsistent with the decision above not to include non-firm generator benefits. It is not really practical to have one rule for reliability expansions and another for economic expansions since, as noted above, it is not always possible to categorise expansions in this way.

A possible mechanism for resolving this dilemma is to incorporate a *contingent auction* into the RIT-T process. This process is outlined below. More detail is provided in appendix F.2.

Where reliability access is provided by an expansion, some firm access could *potentially* be provided instead, using the expanded network capacity. After all, reliability standards are most onerous under peak demand conditions which are similar to FAPS conditions. If access is increased under one set of conditions, it is likely to be increased in the other. Each of the different expansion options being considered in the RIT-T will have the potential to deliver different amounts of firm access, to different generators.

In the contingent auction, a TNSP invites bids from non-firm generators for this potential firm access. The value of the bids is assessed for each expansion option and incorporated into the RIT-T assessment. The preferred expansion option is then chosen and the winning bidders *associated with that option* are awarded firm access and would pay an access charge consistent with their bid. In this way, the auction clearing is *contingent* on the preferred option. It is not the highest overall bid that is cleared in the auction, but the highest of the bids relating to the preferred option.[221]

220 Although their expertise should improve in this area under OFA, in order to manage their market exposure under the incentive scheme.

221 A generator could, of course, "hedge its bets", by submitting bids for all of the options that had the potential to provide it with firm access.

The auction payment will be at a discount to the regulated access price, and so benefits the winning generators. But, nevertheless, it helps to fund at least *some* of the expansion cost and so benefits TUOS customers also. Importantly, because the value of the provided firm access is incorporated into the RIT-T, the efficiency of the decision-making is improved.

On the other hand, this process means that reliability standards are indirectly leading to access price discounting. This outcome was specifically avoided in the design of access pricing refer section 6.2.1, so there is some inconsistency here. The merits of the process depend on the extent to which the contingent auction process impacts on generator procurement behaviour overall: ie, whether it encourages generators to hold off on procurement in the hope of receiving discounted access through the contingent auction. Given the uncertainty around this, the contingent auction is not proposed as a core element of the OFA design.

Future Benefits

Under the RIT-T, a TNSP must assess *future* benefits as well as immediate ones. For example, it may have the choice of two sizes of expansion, both of which are feasible. Although the larger expansion will have a higher cost, it is also likely to provide greater future benefits by continuing to maintain reliability for longer: ie, by pushing back the date of the next reliability expansion. This assessment would be predicated on the demand forecast.

Under OFA, future benefits would also relate to the future ability to maintain FAPS. This assessment would correspondingly be based on forecast levels of firm access. It would be expected that this forecast would be similar to the forecast developed by the AER for the access pricing baseline.

8.3 Design Issues and Options

There are no further design issues and options to consider for this chapter.

9 Transitional Access

9.1 Overview

Transitional Access (TA) would be issued at the time of OFA commencement, with the objectives of mitigating the impact of its introduction and allowing affected parties time to develop their capabilities for operating in the new regime without being exposed to undue risks in the initial period.

TA acts similarly to other firm access, although it has somewhat different terms in relation to renewal rights and sellback rights. However, its shape is different, being a trapezoid with constant volume for the first five years, followed by linear reduction to zero over the next 10 years. The TA term commences at the same time as OFA.

TA issuance differs from other OFA issuance processes, and takes place over two stages. TA issued in the two stages has identical terms and conditions, as described above.

In the first stage, some TA is allocated to existing generators for free, in proportion to their generation capacity. The scaling factor applied will be designed to ensure that, firstly, the issued TA complies with FAPS constraints and, secondly, it does not in aggregate exceed regional peak demand in any region.

In the second stage, an auction is held in which any remaining spare flowgate capacity in the base year is offered at a zero reserve price and in which generators can offer to sell some of their first stage allocation. Bids for both intra-regional TA and transitional FIRs (having terms and conditions similar to TA) could be submitted into the auction.

Apart from the different product being issued, the TA auction will operate in accordance with the same principles as the short-term auction, although settlement might be somewhat different.

To avoid the issued TA prompting network expansion, the introduction of FAPS obligations on TNSPs will be delayed for five years, meaning that shortfalls – under FAPS conditions – will be permitted over this period.

9.2 Design Blueprint

9.2.1 Objectives

The objectives of transitional access issuance are:

- to mitigate any sudden changes to prices or margins for market participants (generators and retailers) on commencement of the OFA regime;
- to encourage and permit generators – existing and new – to acquire and hold the levels of firm access that they would choose to pay for;

- to give time for generators and TNSPs to develop their internal capabilities to operate new or changed processes in the OFA regime without incurring undue operational or financial risks during the learning period; and
- to prevent abrupt changes in aggregate levels of firm access that could create dysfunctional behaviour or outcomes in access procurement or pricing.

Importantly, TA should not delay or dilute the efficiency benefits that the OFA model is designed to promote.

9.2.2 Terms and Conditions

Overview

The terms of TA are non-standard in the following ways:

1. All TA is issued with the same shape which is a *trapezoid* rather than the usual rectangular (strip) shape.
2. TA has *renewal rights* with a specified renewal term.
3. TA does *not* have sellback rights.

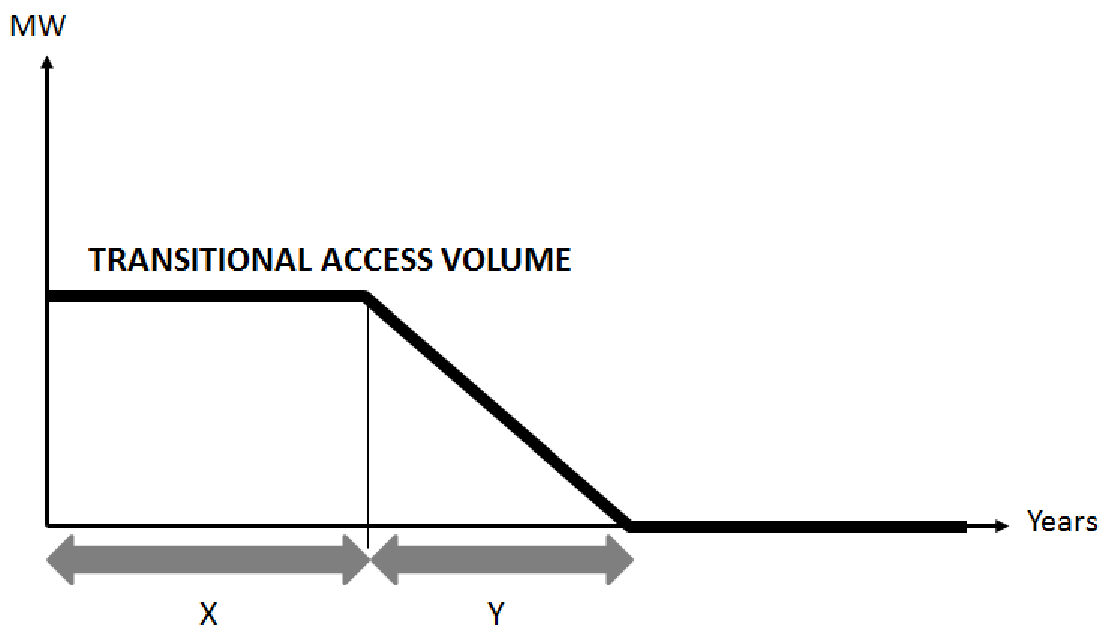
These differences are detailed in turn in the sections below.

Aside from the differences listed above, TA operates identically to FA under the OFA model, ie:

1. Details of TA are recorded in the firm access register.
2. TA is awarded entitlements in access settlement.
3. TA creates obligations for TNSPs under FAS, although the introduction of FAPS obligations will be delayed.

Sculpting

All issued TA has the same trapezoid shape, presented in the figure below.



The TA amount is:

- constant for X years; and
- declines linearly to zero over the next Y years.

It is proposed that:

- X is set equal to 5 years; and
- Y is set equal to 10 years.

Issues around setting the *sculpting parameters* (X and Y) are discussed in section 9.3.3.

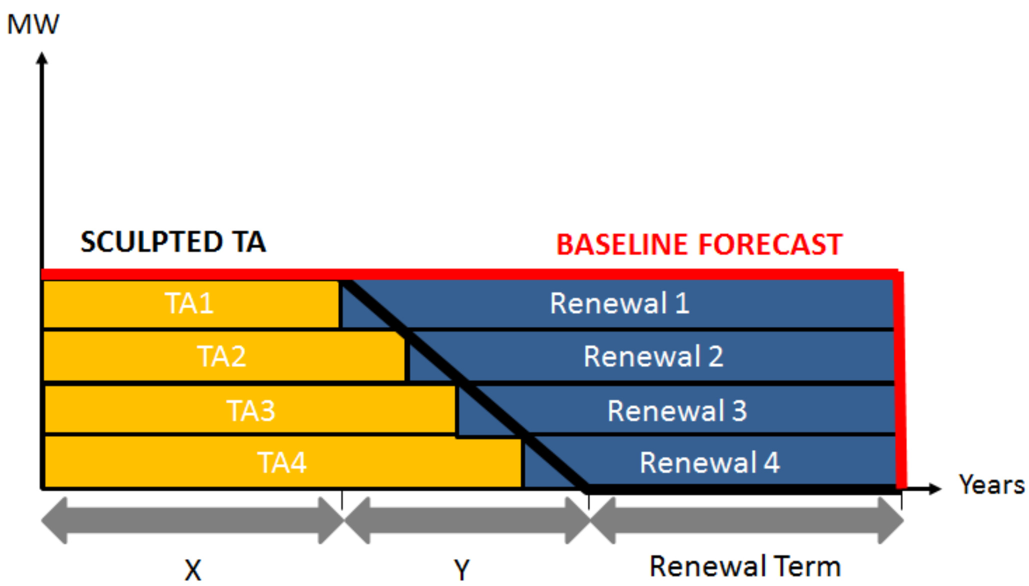
Renewal Rights

Renewal rights were defined in section 6.2.2. Firm access with renewal rights has its renewal *anticipated* in the baseline forecast. For reasons discussed in section 6.2.3, this means that a renewal request (ie, a request that seeks to renew some firm access that has renewal rights) is priced at LRDC, rather than the usual LRIC that is charged to other, non-renewal requests.

All TA is provided with renewal rights. The renewal term will be specified at the time that the TA is issued. It will be set to reflect expected the generator asset life: ie, the *end* of the renewal period will coincide with the expected *end* of the asset life. This assumption is used only for the baseline *forecast*: there is no *obligation* on generators to renew for exactly this period, or even to renew at all.

The implied baseline forecast for firm access is presented in Figure 9.1 (which for simplicity this does not show any modelled firm access). For illustrative simplicity, the example has Y=4 rather than Y=10. The inclusion of the anticipated renewals means that the baseline forecast, has a rectangular shape, rather than the trapezoidal TA shape.

Figure 9.1 TA baseline forecast



The figure illustrates how the trapezoid shape of TA can be considered to consist of several strips of different terms. Each strip would have its own renewal right. In the figure, each strip has a renewal right with a different term, in order that the renewals all end at the same time. Alternatively, a common renewal term could be used, in which case the baseline forecast would have a trapezoid shape.

The figure shows what the baseline forecast would look like at OFA *commencement*. During the Y period,[222] generators would need to decide whether or not to renew TA, and for what amount and term. Therefore, the anticipated renewals in the baseline forecast would progressively be replaced with the actual renewals, or with nothing, if there is no renewal.

Renewal rights on TA are discussed further in section 9.3.3.

222 In fact, starting three years before the Y period, because of the transmission planning lead time.

Sellback Rights

As described in section 7.2.2, *purchased* long-term intra-regional firm access can be sold back to the TNSP, at any time,[223] at the *current* LRDC price.

Sellback of transitional access, on the other hand, is not permitted. The reasons for this decision are discussed in section 9.3.3.

FAPS Amnesty

FAPS obligations on TNSPs would not be introduced until the end of the X period. That is to say, a TNSP would not be required to plan to remove any shortfalls that might arise – under FAPS conditions - *during* the X period. It would, nevertheless, need to undertake transmission planning during the X period in order to address possible shortfalls arising in the Y period and beyond. This is discussed further in section 9.3.4. Notwithstanding the amnesty, FAPS shortfalls during the X period are likely to be minimal since TA issuance will be calibrated to the FAPS capacity existing at the commencement of OFA.

9.2.3 Issuance

Overview

TA issuance takes place at the time of OFA implementation. Issuance takes place in two stages:

- *TA allocation*: a free allocation to existing generators, in proportion to their *existing capacity*; and
- *TA auction*: an auction of any remaining spare capacity.[224]

The terms and conditions of the TA issued in each stage would be as described in the previous section.

TA Allocation

In the first stage, TA is allocated to generators in proportion to their *existing capacity* by applying a *regional scaling factor* to generators in each region, ie:

TA allocation = regional scaling factor × existing capacity

Existing capacity needs to be defined carefully, in accordance with the transition objectives. This is discussed further in section 9.3.1.

---

223 Subject to some sellback restrictions described in section 7.3.9.

224 Spare capacity here has the same meaning as under the short-term auction: ie, FAPS capacity minus TFGX.

The scaling factor is set at the highest value consistent with the following requirements:

1. aggregate TA in a region does not exceed regional peak demand; and

2. allocated TA complies with FAPS requirements: ie, the implied TFGX does not exceeds FAPS capacity.

These requirements would be applied only in relation to the TA base year: ie, the first year of OFA operation.[225]

To meet the first requirement, the regional scaling factor can be no higher than the peak demand factor, defined as:

> Peak demand factor = regional peak demand / total existing capacity in the region

To meet the second requirement, on a particular flowgate, the regional scaling factor can be no higher than the flowgate capacity factor, defined as:

> Flowgate capacity factor = FAPS capacity / unscaled TFGX

Where the *unscaled TFGX* is what the TFGX would be if all generators were allocated TA *equal* to their existing capacity. There will be a different flowgate capacity factor for each flowgate.

The regional scaling factor would be set at the *minimum* of all of these factors:

> Regional scaling factor = min(peak demand factor, lowest flowgate capacity factor)

TA Auction

In the second stage, an auction of TA takes place. The auction formulation and objectives would be similar to those of the short-term auction, as described in section 7.2.4. In particular:

1. The total TA issuance (from both stages) must comply with the FAPS requirement: this requirement applies only to the TA base year.[226]

2. Market participants can bid for FIRs: these will be "TA FIRs" in the sense that they have the TA shape and terms.

3. Generators can offer all or part of their allocated TA for sale in the auction.

4. Auction participants are not limited in how much they can bid for: eg, they could bid to become super-firm.[227]

---

225 Although it would be fairly straightforward to apply them to all years in the X period.

226 This means that the FAPS constraints could potentially be breached in subsequent years. This is discussed in section 9.3.4.

5. Auction revenue is allocated to TNSPs, in accordance with flowgate tagging.

6. Revenue regulation is defined to pass-through auction revenue to TUOS customers.

There are some important differences between the TA auction and ST auction:

- The term and shape of the issued TA would be different to the FA issued in a short-term auction.
- TNSPs would not be permitted to participate: eg, by making buyback bids or by offering discretionary capacity.
- Settlement could be by instalment, governed by payment deeds established between buyers and TNSPs. This is discussed in the next section.

Auction Settlement

The TA auction could be settled in two ways:

- cash-settled, in the same way as the short-term auction; and
- settled in instalments, similar to the long-term inter-regional auction.

TA is long-term and it would be consistent with the treatment of other long-term purchases to settle it through instalments. For this, a payment deed would be established between buyers and TNSPs, defining the payment profile as well as prudential requirements and termination provisions.

On the other hand, long-term sellbacks - in the context of long-term intra-regional and inter-regional issuance - are settled *immediately*, which might suggest that sellers into the TA auction should be paid immediately also. However, this would create a cash shortfall for the TNSP: with sellers being paid immediately, but buyers paying the TNSP only in instalments. To avoid this, there should be a *consistent* approach to settling buyers and sellers.

The magnitude of clearing prices and revenues in the TA auction are unclear at this stage. If prices are likely to be fairly low, cash settlement is preferable, since it avoids the need to establish payment deeds. If they are likely to be high, instalments may be needed. For this reason, the design decision should be made closer to OFA implementation, when TA prices become clearer.

227 This is discussed in section 9.3.6.

9.3 Design Issues and Options

9.3.1 Existing Capacity

TA allocation is based on a definition of the existing capacity of generators. In access settlement, generator capacity is defined as the maximum output of the generator over the previous 2 years, or other defined recent historical period. However, using this definition may be inappropriate in the context of TA allocation since it would mean:

1. Generators that had been mothballed, or had otherwise not operated, for at least two years would *not* be included.

2. Generators that had previously operated within the last two years, but were now closed *would* be included.

3. New generation developments that were committed but had not yet commenced full operations would be partially or fully *excluded.*

The treatment of these generator categories in the context of TA issuance needs some further consideration, informed by the transition objectives.

9.3.2 Setting the Sculpting Parameters

Overview

The choice of sculpting parameters is informed by the transition objectives. For convenience, these are repeated below.

- *mitigate impacts*: to mitigate any sudden changes to prices or margins for market participants (generators and retailers) on commencement of the OFA regime;

- *efficient firmness*: to encourage and permit generators – existing and new – to acquire and hold the levels of firm access that they would choose to pay for;

- *learning period*: to give time for generators and TNSPs to develop their internal capabilities to operate new or changed processes in the OFA regime without incurring undue operational or financial risks during the learning period; and

- *orderly pricing and procurement*: to prevent abrupt changes in aggregate levels of agreed access that could create dysfunctional behaviour or outcomes in access procurement or pricing.

The implications of these objectives for the sculpting parameters are considered in turn below.

Mitigate impact of OFA commencement

The impact of OFA commencement will primarily arise from three factors:

- the change in the level, and certainty, of access enjoyed by a generator under OFA, compared to the *counterfactual* of a continuation of current arrangements;
- the access charges that a generator pays under OFA; and
- any changes in regional prices (spot or forward) resulting from OFA implementation.

Access under current arrangements is based on dispatch. Aggregate access will depend upon flowgate capacity which, in turn, depends upon network capacity. Network expansion is primarily driven by reliability standards, which require aggregate access to be expanded in tandem with growth in peak demand. On the other hand, demand growth will attract new entrants, with whom existing generators would have to share the aggregate access. Taking these various offsetting factors into account, the aggregate access enjoyed by *existing generators*[228] under the counterfactual is unlikely to change substantially over the medium term. On an individual basis, of course, access levels may change. In particular, generators "lose" access when they close, because they are no longer dispatched.[229]

Under OFA, generators will have to pay access charges to:

- purchase TA in the auction; and
- renew TA when it is sculpted back below their desired access level.

The cost of TA purchases is unclear. It could plausibly be argued that:

- generators require TA to back their forward contracts;
- the aggregate volume of forward contracts should broadly reflect peak demand; and
- therefore, the aggregate TA requirement should also reflect peak demand.

In this case, if there is sufficient FAPS capacity so that the aggregate allocated TA equals peak demand, generators' aggregate access requirements will largely be met. There will be some additional purchases – there must be if there is some spare capacity available to be sold in the auction – but the prices for those purchases might broadly reflect "fair value": ie, what the purchaser can expect to get back through access settlements. To summarise, the cost of TA purchases could plausibly be relatively low.

228 Ie, those that would be allocated TA at OFA commencement.

229 Or, in fact, two years after they close, because of the way that generation "capacity" is defined.

The impact of OFA introduction on regional prices is uncertain. In the long-run, regional prices should, for reasons presented below, reflect and lie somewhere between the:

1. *entry cost* for new generation; or
2. *exit value* for existing generators.

New generators will enter if the regional price consistently exceeds their *entry cost*. Their entry will then cause the price to fall. Therefore, the regional price is unlikely to exceed entry cost over the long-term, assuming that the market is contestable.

On the other hand, if the regional price is below the *exit value* – the costs that a generator can avoid by closing – then a generator will close and this will cause the price to rise. Therefore, the regional price is unlikely to be below exit value over the long-term, assuming that there are no barriers to exit.

The impact of OFA depends upon the preferred firmness of new and existing generators. The preferred level is uncertain, but it seems reasonable to assume that it is broadly the same for new and existing generators.

On this assumption, the associated access charge is a component of *both* the entry cost and the exit value: it must be purchased by new entrants; and it is no longer required by a generator who exits. Therefore, based on the analysis above, one would expect the regional price to adjust – over the long-term – to reflect this access charge. Of course, the access charge is different for each generator, so the regional price might reflect a typical or average access price. Nevertheless, the additional revenue to generators from this price adjustment will, over the long-run, broadly reflect and offset access prices.

To summarise these various arguments:

- The aggregate access of existing generators, under the counterfactual, should not change substantially over the medium-term and will reflect peak demand.
- Assuming sufficient existing FAPS capacity, generators will – in aggregate - be allocated a level of TA similar to their existing access.
- The price and cost of TA purchases will be relatively low, reflecting expected congestion costs over the TA period.
- Generators will incur access charges in replacing their TA when it is sculpted back.
- Over the long-run, regional prices will reflect the cost of purchasing this level of firm access.

This suggests that TA should achieve its objective of mitigating the impact of OFA on existing generators so long as the timescale of the sculpting is similar to the timescale over which regional prices adjust to accommodate and reflect the cost of purchasing FA. This latter timescale is unclear, but could perhaps be estimated through some quantitative modelling of future generator entry and exit, and associated price changes, under OFA.

Efficient Firmness

A key element of OFA is that it provides *price signals* to generators in relation to location and firmness decisions. So long as generators are responsive to these signals, the co-ordination and efficiency of generation-transmission planning should be improved, as discussed in section 6.3.3.

This transition objective reflects a concern that, if generators are gifted access through TA, they might not respond to access prices: ie, they might continue to hold the TA, despite the fact that, at current prices, they would not have chosen to buy that access.

If there is an efficient secondary market, rational generators should respond to prices in that market, irrespective of how much TA they are allocated. In principle, the short-term auction platform should facilitate an efficient secondary market. However, the long-term secondary market may be less efficient, particularly when it is a new entrant generator who is seeking to purchase firm access from another firm generator, for the following reasons:

- an existing generator may be unwilling, for strategic reasons, to facilitate the entry of a new generator by selling firm access to it;
- the term of the TA – which the generator is able to sell – might not meet the requirements of the entrant, who is likely to seek to buy long-term FA to match its expected asset life; and
- TA does not have sellback rights, and so cannot be sold into a secondary trade. Although, possibly, sellback rights could be provided in this case.

In summary, this objective is likely to be best achieved by sculpting back TA over the medium-term.

Learning Period

Operational and financial risks arise if generators are forced to buy or sell firm access – and also if TNSPs are required to issue it – during the learning period when the associated capabilities (pricing models, decision-support processes, etc) are not fully developed.

This "forced" purchase is likely to be associated with a generator having insufficient firm access to back its forward position which, in turn, is only like to occur during the Y period.[230] Therefore, X needs to be long enough to cover the learning period.

230 As discussed previously, generators are likely to be allocated sufficient TA in this respect initially.

A learning period of one or two years may be sufficient. But recall that long-term firm access must be purchased with a three year lead time. This implies that X should be five or more years.

On the other hand, *discretionary* FA trading should be encouraged during this period, so that there is "learning by doing". The TA auction helps to do this, by providing participants with an opportunity to trade at a relatively low risk.[231]

Orderly Pricing and Procurement

As TA expires, generators are likely to procure new firm access to replace it. Since TA has renewal rights, this involves generators making renewal requests.

As discussed in section 7.3.2, TA renewals are likely to present the major challenge to the long-term intra-regional issuance process and the associated queuing policy. Since all existing generators will have TA, and many will be submitting renewal requests as their TA expires, there will be many concurrent requests which need to be queued.

The TA shape – with volume tapered over the Y period - is designed to mitigate these problems. For example, if Y=10, as proposed, then only 10% of TA will expire in each year during the Y period. If there is an annual season, each annual season will then need to handle at most 10% of the TA volume. The longer the Y, the lower the annual MW *volume* of TA renewals: although, the *number* of annual renewals will be largely unaffected by Y.

As discussed previously, because renewal rights are provided on TA, the baseline forecast will include anticipated renewals of all TA initially, but will progressively replace this with actual renewals (or nothing, if renewals are not taken up) over the Y period. The *longer* that Y is, the more *gradual* the baseline firm access shape changes and this should help support this transition objective.

In summary, this objective is primarily related to Y: the longer the Y, the better the objective is achieved.

Summary

The relevance of the four transition objectives for the setting of X and Y are summarised in Table 9.1 below.

[231] In this respect, it might be desirable for there to be several TA auctions. These could progressively sell any spare FAPS capacity, similar to how the short-term auction operates.

Table 9.1 Relevance of transitional objectives for setting X and Y

Objective	Relevance for X	Relevance for Y
Mitigate impact	Match price adjustment	Match price adjustment
Respond to price signals	Not too long	Not too long
Learning period	At least five years	Not relevant
Avoid dysfunctional procurement	Not relevant	The longer the better

The proposed settings of X=5 and Y=10 are therefore broadly consistent with the objectives. The objectives conflict for both X and Y and the chosen settings appear to be a reasonable trade off. However, some more detailed analysis – in particular in relation to the impact of OFA on regional prices – may suggest some alternative timings.

9.3.3 TA Terms

Renewal Rights

As was discussed in section 6.3.4, there are two conflicting objectives in relation to renewal rights:

- to ensure that access prices are unbiased; and
- to promote competition by ensuring that the price of an access request does not depend upon who is requesting it – all other things being equal.

It was argued there that these objectives are best achieved if purchased firm access is not allocated renewal rights. This was on the basis that:

- the term of purchased firm access is likely to reflect expected asset life, so a renewal of this firm access should *not* be anticipated in the baseline forecast; and
- the financial advantage from renewal rights held by *existing* generators could create a barrier to entry to *new* generators.

These arguments do *not* apply to transitional access since the X + Y term of TA is *not* designed to reflect asset life and so it would be appropriate to anticipate renewal in the baseline, at least for long-life generators.

As discussed, the X and Y values will be set with the objective, amongst other things, of mitigating the impact of OFA on existing generators. Consistent with this objective, any beneficial effect of renewal rights could be offset by an adjustment to the X and Y values.

No Sellback Rights

In the light of the transition objectives, there are a number of reasons why sellback rights should not be provided on TA:

1. TA is designed to provide a stable pricing baseline for access pricing during the transitional period. Significant sell backs of TA could disrupt this baseline.

2. The access pricing model may lack robustness initially and could misprice sellback requests.

3. Policy concerns arise at the possibility that TA that has been gifted to existing generators could subsequently be profitably sold back to TNSPs.

On the other hand, as noted above, providing sellback rights could improve long-term secondary trading and so help to achieve the objective relating to price signals. It would be feasible to permit sellback rights to facilitate a secondary trade[232] without conflicting with any of the reasons listed above.

Secondary Trading

Effective secondary trading of access requires all firm access to have common terms. If there were many different types of firm access, the secondary market would become complex and fragmented. However, because of its special terms, TA is a slightly different form of firm access and this could potentially impact on secondary trading.

There should be *no* impact on short-term secondary trading (through the *short-term* auction) since, by definition, short-term firm access does not have any sellback or renewals rights anyway. These both relate to the long-term procurement process.

Secondary trading of TA can occur through the TA auction. In this case, TA is bought and sold and so the different terms of TA do not matter.

Secondary trading of TA in the long-term is likely to be limited, for reasons discussed above. Were it to be permitted – by providing sellback rights for this purpose – it would be important to ensure that the FA that was purchased in the trade had the same terms and conditions as the sold TA: in particular, that it also did not have sellback rights.

9.3.4 FAPS Amnesty

TA could trigger expansion

TA allocation must comply with FAPS constraints in the *base year*. However, if flowgate capacity subsequently reduces – at a rate faster than the TA is sculpted back – there

232 Ie, where the request was part of a combined sellback-purchase request as described in section 7.3.6.

could be breach of FAPS constraints later during the TA term. A possible example of this is where a network asset reaches the end of its life during the TA term. Ordinarily, under FAPS, the TNSP would be required to expand or replace network capacity to address the breach. This risk could arise during the X period and could possibly continue into the Y period, until the sculpting back of TA removed any possible FAPS breaches.

Delayed FAPS introduction

This issue is addressed by *delaying* the introduction of the FAPS obligation until after the X period. That could mean that TA (and any purchased FA) is slightly less firm over this period.

Given that a TNSPs planning lead time is at least 3 years (compared to a proposed X of 5 years), the FAPS delay will only potentially impact over a relatively short period.

A corresponding adjustment to the incentive scheme may be appropriate. This could be done by delaying introduction of the scheme until the FAPS is introduced, or by adjusting the benchmark shortfall cost to reflect any FAPS breaches over this period.

9.3.5 Interconnectors

No Inter-regional TA

Interconnectors have lower firmness of access under current arrangements, because they are unable to compete with generators for dispatch under race-to-the-floor bidding conditions. Consistent with the "mitigate impact" transition objective, the TA allocation reflects the existing situation, irrespective of its merits.[233] If TA were to be allocated to interconnectors, this would mean less TA for generators, who could suffer a bigger impact from OFA's introduction as a result.

Inter-regional Reservation

Suppose it were decided, notwithstanding the above discussion, to allocate some firm access to interconnectors during the transition period. How would this be done and what would this mean for OFA outcomes?

TA could be allocated to DICs, by including them in stage 1 of the TA allocation process. But it is not clear what would then be done with this allocated TA: DICs are *notional* entities, so the TA must then somehow be re-allocated to entities that are *real*. Intra-regional TA is allocated to those parties (ie, existing generators) who would freely receive access if the status quo continued. But inter-regional access is not

---

233 Arguably access for interconnectors is *improved* under OFA, even with the proposed TA allocation, because interconnectors at least get *zero* access, whereas currently they commonly have *negative* access: ie, during counter-price flows.

allocated freely to *any* market participants under the current arrangements; it must be *purchased* through the settlement residue auction.

To reflect this, the inter-regional TA would also need to be auctioned. This could be through:

- the TA auction; or
- a new and separate inter-regional TA auction.

If it were offered into the TA auction, it would need to have a zero reserve price to ensure that it was sold. The proceeds from this sale could then be passed to TUOS customers, along with the rest of the TA auction proceeds.

But, if the TA auction is efficient, the pre-auction TA allocation should be irrelevant to the post-auction TA holdings. So the fact that DICs were allocated more TA, and generators less, will not change how much TA each party *eventually* ends up with. It simply means that generators will have purchased more TA – or sold less - at the auction. The higher generator payments flow to the DICs selling the allocated TA and, in turn, to TUOS customers.

In summary, the inter-regional reservation will not change how TA is held, post-auction. It will simply lead to a new *wealth transfer* from generators to TUOS customers. If this were considered desirable (and it is not clear why it would be), it could be effected more easily through an adjustment to the sculpting parameters.

Alternatively, if the inter-regional TA was sold, through a separate inter-regional auction process, the allocated inter-regional TA would be translated into FIRs. But if, in fact, providing *intra*-regional access (on interconnector paths) instead of inter-regional access is a more valuable use of inter-regional flowgate capacity, this *forced* allocation to FIRs is not an efficient outcome. If, on the other hand, FIRs *are* more highly valued, they would have been purchased in the TA auction anyway, without the need for any inter-regional reservation.

Summary

In summary, there are several reasons for not allocating TA on interconnectors:

- Zero inter-regional TA reflects interconnector access under current arrangements.
- Inter-regional TA cannot be held by DICs (who are not real entities); it would have to be sold onto market participants through an auction process.
- If it were sold through the TA auction, the outcome would be no different from the zero TA approach, except for a wealth transfer from generators to TUOS customers.

- If it were sold as long-term FIRs, this could inefficiently lock the associated flowgate capacity into inter-regional firm access provision, when it may be more highly valued when providing intra-regional access.

9.3.6 Super-firm Auction Purchases

Overview

It is proposed that there is no restriction on purchases of TA in the TA auction. In particular, there is no restriction on generators becoming super-firm through the TA auction.

For example, a 1,000MW generator who was allocated 700MW in the first TA issuance stage would be permitted to then purchase a further 600MW in the TA auction. As a result, the generator would be super-firm: holding 1,300MW of TA, higher than its 1,000MW capacity.[234]

The clearing price for this purchase might be low, if surplus existing transmission capacity means that there is limited congestion affecting that generator currently.

Concerns

There is a possible concern that generators may be able to *squat* on the existing transmission capacity: buying it up very cheaply in the auction and then holding onto it in order either to deter, or to profitably sell on to, future new entrants.

These concerns are mitigated by the sculpting of TA. Spare capacity will become available anyway during the Y period, as the TA is scaled back.

In any case, if super-firm purchases were *prohibited*, the auction clearing prices would be commensurately *lower*. There could even be a situation where not all of the FAPS capacity is sold, even with all generators buying to cover their full capacity, in which case the clearing price would be *zero*. In this case, generators genuinely *are* squatting: the cure is potentially *worse* than the disease.

Possibly, future entrants might wish to bid into the auction. By definition, any purchases would make them super-firm, since they have zero existing capacity. Prohibiting super-firm purchases would have the effect of barring future entrants from the auction. This adds to, rather than removes, entry barriers.

Conclusion

Although generators might be able to effectively squat on existing capacity if they are permitted to become super-firm with TA, simply prohibiting super-firm purchases is

234 The generator will not obtain a super-firm service, though, as a result. It gains no extra payments from access settlements for the extra 300MW of TA.

likely to be counterproductive, by reducing auction prices and barring new entrants from participating.

Instead, any concerns about entry barriers being created as a result should be addressed through the sculpting of TA.

A Access Settlement Concepts

A.1 Flowgate Pricing

A.1.1 Overview

There is something of a disconnect between the high level description of access settlement (set out in chapter 2) and the more detailed design (described in chapter 4). The former description is based on local prices (LMPs) and the latter description uses flowgate prices. It is noted that, in the absence of any scaling back of entitlements, the two approaches are mathematically equivalent.

However, that assertion is far from obvious or intuitive. This section explains the relationship between flowgate prices and local prices and demonstrates that the two approaches can be equivalent. Most importantly, it demonstrates that the settlement algebra successfully achieves the *no-regret dispatch* principle: that all dispatched generators are paid, net, at least their offer price for their output, irrespective of their level of access or entitlements.

A.1.2 Transmission Constraint Equations

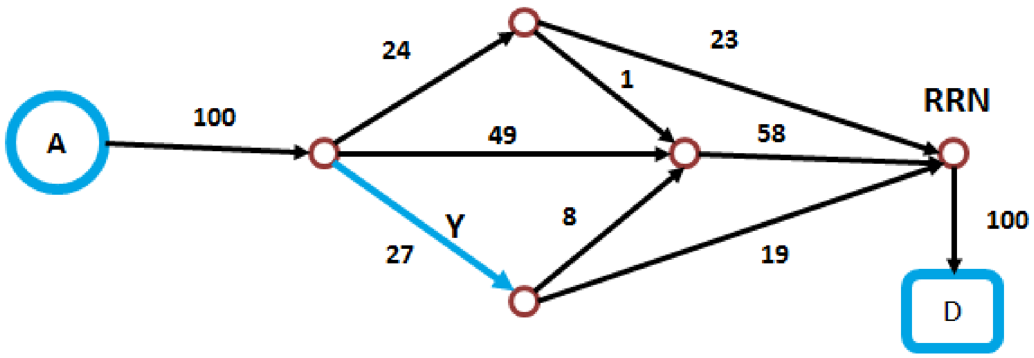
Participation Factors

Load flow analysis on a meshed AC power system is extremely complex. It can be considerably simplified by making two approximations:

- *DC approximation*: MVAr flows and variations in voltage magnitudes are ignored; and
- *lossless approximation*: transmission losses are zero.

With these approximations – which are not material for the purposes of this discussion – AC load flow is similar to the electricity flow through a DC circuit or a liquid flow through a network of pipes. Figure A.1 below, illustrates a transfer of power through a network, from a generator node to the RRN, based on the lossless DC approximation.

Figure A.1 Load Flow Example 1

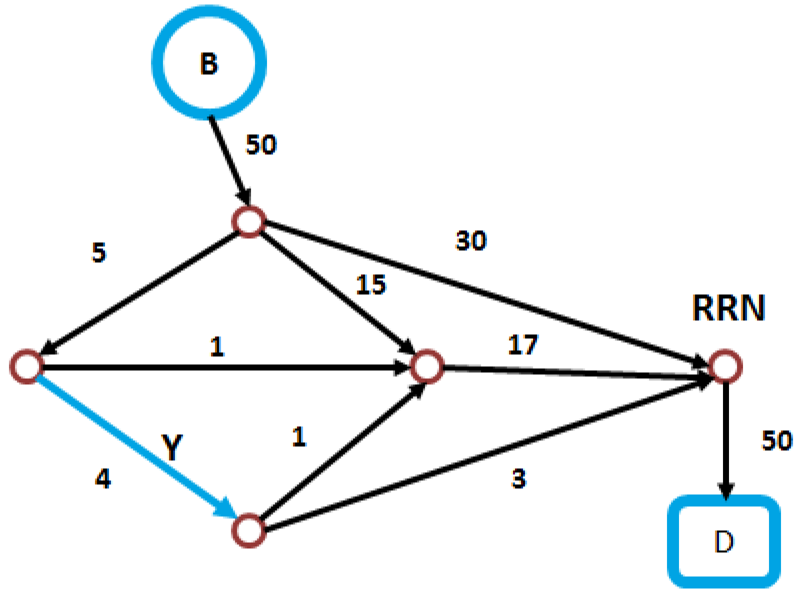


Each branch in the network has the *potential* to become congested – where the flow on the line reaches the line's maximum rating - and so can be regarded as a *flowgate*, in the terminology of the OFA model. One of these flowgates is marked on the figure, labelled *flowgate Y*. Of the 100MW injected by Generator A, 27MW flows through flowgate Y. The *participation* of generator A in flowgate Y is therefore 27 per cent. Note that both the flowgate and the flow have a specified direction. Because they are in the same direction, the participation factor is *positive*; if they were in opposite directions the participation factor would be *negative*. Every branch can, in principle, be congested in either direction, meaning there are two (directed) flowgates on each line.

Note that the analysis is *oriented* to the RRN: the output from Generator A is assumed to be supplying a load of equal size that is connected at the RRN. Because of its fundamental role in setting regional prices, participation factors are *always* oriented to the RRN.

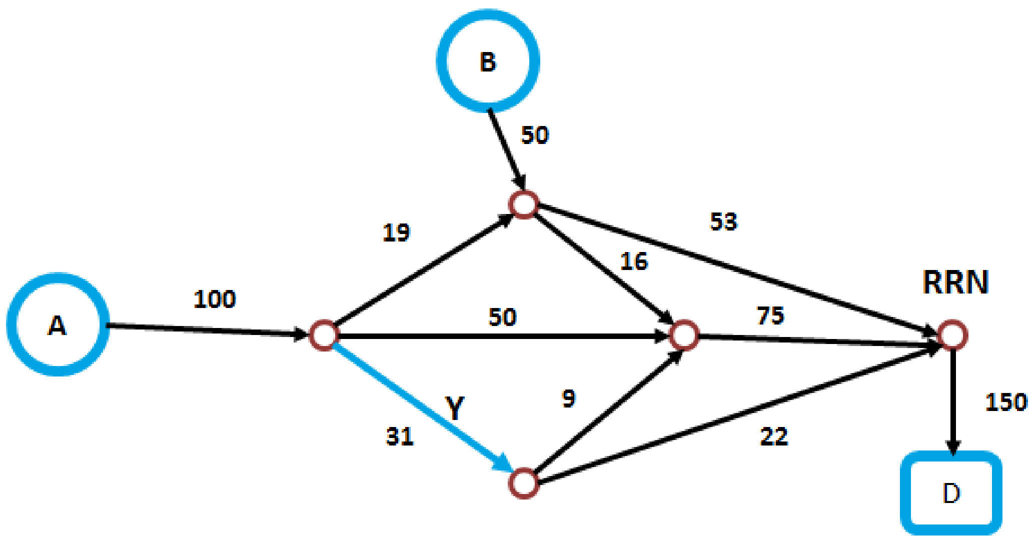
A similar load flow for generator B is presented in Figure A.2 below. Because the generator connects at a different node, it has a different participation factor. The participation of generator B in flowgate Y is 4 MW divided by 50 MW, or 8 per cent.

Figure A.2 **Load Flow Example 2**



Our simple model of load flow is *linear*: flows are proportional to injections so if the injection is doubled, say, the flow on each branch would double. Linear systems permit *superposition*: adding together two load flows creates a third load flow. If two the load flows presented above are superimposed, a third load flow is created as shown in Figure A.3.

Figure A.3 **Superimposed Load Flow**



It is seen that the flow through flowgate Y is the sum of the flows in the previous two load flows, ie:

Flowgate Y flow = 31 = 27 + 4 = 27% x 100 + 8% × 50

In general, the flow on the flowgate can be calculated using the formula:

$$\text{Flowgate Y Flow} = 27\% \times G_A + 8\% \times G_B \text{ (A1.1)}$$

Where:

G_A is the output of generator A

G_B is the output of generator B

27% is the participation of generator A in flowgate Y

8% is the participation of generator A in flowgate Y

If the limit on the line associated with Flowgate Y is TX_Y then, for any dispatch of generators A and B, we must have:

$$\text{Flowgate Y flow} = 27\% \times G_A + 8\% \text{ x } G_B \leq TX_Y \qquad \text{(A1.2)}$$

This inequality has the form of the transmission constraints used in NEMDE.

The general version of this inequality is:

$$\sum_i \alpha_{ik} \times G_i \leq TX_k \qquad \text{(A1.3)}$$

Where:

α_{ik} is the participation of a generator i in flowgate k

G_i is the output of generator i

TX_k is the transmission capacity of flowgate k

Flowgate Capacity

Suppose that there is some local demand and also some non-scheduled generation (NSG) at nodes A and B. For our simplified model, the impact of non-scheduled and scheduled generation is identical. The impact of demand on network flows is exactly the opposite to generation: ie, it is the *net injection*, G+NSG-D, at each node that determines flows; scheduled generation of 100MW will create the same flow as scheduled generation of 130MW, non-scheduled generation of 10MW and local demand of 40MW, say (since 130 + 10 - 40 = 100).

When local demand and NSG is included, inequality (A1.2) becomes:

$$27\% \times (G_A + NGS_A - D_A) + 8\% \times (G_B + NSG_B - D_B) \leq TX_Y \text{ (A1.4)}$$

Where:

D_A is the local demand at node A

D_B is the local demand at node B

NSG_A is the NSG at node A

NSG_B is the NSG at node B

NEMDE constraint formulation uses the convention that dispatchable variables (ie, dispatch targets for scheduled- and semi-scheduled generation and scheduled demand) are placed on the left-hand side (LHS) of the constraint equation and all other variables (including NSG and demand) are placed on the right-hand side (RHS). Therefore, the demand and NSG in the inequality above is moved to the RHS of the inequality and equation (A1.3) becomes:

$$27\% \times G_A + 8\% \times G_B \leq TX_Y + 27\% \times D_A + 8\% \times D_B - 27\% \times NSG_A - 8\% \times NSG_B \equiv FGX_Y$$

Where:

$$FGX_Y = TX_Y + 27\% \times D_A + 8\% \times D_B - 27\% \times NSG_A - 8\% \times NSG_B \text{ (A1.5)}$$

FGX_Y is the *flowgate capacity* for flowgate Y. It is seen that flowgate capacity is a combination of the transmission capacity, local demand and NSG. The general form of equation (A1.5) is:

$$FGX_k = TX_k + \sum_i \alpha_{ik} \times D_i - \sum_i \alpha_{ik} \times NSG_i$$

Where:

FGX_k is the capacity of flowgate k

TX_k is the transmission capacity of flowgate k

α_{ik} is the participation of node i in flowgate k

D_i is the local demand at node i

NSG_i is the NSG at node i

Defining flowgate capacity in this way, equation A1.4 can be generalised to the form:

$$\sum_i \alpha_{ik} \times G_i \leq FGX_k \text{ (A1.6)}$$

It is seen from the above formula that flowgate capacity will vary as local demand and NSG varies. When TNSPs are managing their networks so as to maintain FAS, they must take account of local demand and NSG as well as transmission availability.

Stability Constraints

The discussion above describes the nature of a *thermal* transmission constraint. It has this name because the limitation on the power flow on the line is essentially to prevent it from *overheating*. The *transmission capacity* referred to above is, in fact, the *thermal rating* of the transmission line associated with the thermal flowgate. A thermal constraint has been used to illustrate the concept of flowgates because its characteristics in a real-life AC power system are similar to those under a simplified DC, lossless approximation, the latter being fairly straightforward and intuitive to understand.

Many transmission constraints are *stability* constraints rather than thermal constraints. Unlike thermal constraints, stability constraints only arise in an AC power system and have no DC analogy. They are extremely complex to understand and analyse, and it would not be helpful or appropriate to try to explain them in this document.

In any case, a detailed explanation of stability constraints is unnecessary. Stability limits are highly non-linear. But because NEMDE can only deal with linear equations, AEMO has to undertake mathematical modelling of the stability limits in order to formulate a *linear approximation*, which has the same LHS form as the in equation A1.6.

The complexity inherent in stability limits is then "buried" in the FGX variable. In formulating constraints, AEMO must work out the complex (and commonly non-linear) expression for calculating FGX. But OFA does not need or use this expression. Rather, it *infers* the flowgate capacity on congested flowgates based on the value of the LHS of the constraint equation, which is much simpler to calculate.

The algebraic commonality between thermal and stability constraints – at least as far as access settlement is concerned - mean that all of the analysis in the remainder of this section applies equally to thermal and to stability constraints. Only the illustrations are not applicable to stability constraints since, unlike with thermal constraints, these cannot be considered to be located on a particular line but rather exist more nebulously across the network as a whole.

A.1.3 Flowgate and Local Pricing

Flowgate Prices

In the OFA model, the *flowgate price* is defined as *the marginal value of flowgate capacity for economic dispatch.*

Economic dispatch is the dispatch of generation that meets demand, complies with transmission constraints and minimises generation costs, as these are specified in dispatch offer prices.

The *marginal value* of flowgate capacity is the *increase* in economic dispatch cost caused by a 1MW *decrease* in flowgate capacity.[235] Therefore:

$$FGP = c(ED_2) - c(ED_1)$$

[235] Or, alternatively, the decrease in economic dispatch cost caused by a 1MW increase in flowgate capacity. These two values are generally the same, and the situations where they are different are not important to this discussion.

Where:

ED_1 is the original economic dispatch

ED_2 is the revised dispatch when the flowgate capacity is reduced by 1MW

c(ED) is the cost of an economic dispatch

Characteristics of flowgate prices are:

- *a flowgate price can never be negative*: $c(ED_2)$ cannot be less than $c(ED_1)$. If it were, ED_2 would have been preferred to ED_1 in the original dispatch and so ED_1 would not be economic;
- *if a flowgate is uncongested, its price is zero*: since there is already some unused capacity on the flowgate, removing 1MW of this unused capacity is not going to affect dispatch: ie, ED_2 is the same as ED_1; and
- *if a flowgate is congested then its price is greater than zero*: if the economic dispatch fully uses flowgate capacity, it is no longer feasible when a 1MW of flowgate capacity is removed, and so a more expensive economic dispatch must replace it.[236]

Locational Marginal Price

The locational marginal price (LMP) at a node is defined as *the marginal value of generation at a node*. As discussed previously, the impact of +1MW of generation within a region is identical to -1MW of demand at that RRN. Therefore, the local price is also the *marginal cost of supplying demand at a node*.[237]

Using the same definition of marginal value as above, marginal value is also the amount by which the cost of economic dispatch *reduces* if a zero cost generator, at the node, and not included in the original economic dispatch, injects 1MW.

Suppose that the LMP at a node A is P_A. If the extra 1MW generated at node A is zero cost, the dispatch cost saving is P_A, by definition. More generally, if the 1MW generation costs C_A, the dispatch cost saving is P_A-C_A. If the generator at node A is available for dispatch and submits a dispatch offer price C_A then:

- if $C_A<P_A$, dispatching that generator by 1MW reduces the cost of dispatch: hence that 1MW will be *included* in an economic dispatch; and
- if $C_A>P_A$, dispatching that generator by 1MW *increases* the cost of dispatch: hence that 1MW will *not* be included in an economic dispatch.

236 There may be some special circumstances where alternative economic dispatches, with the same cost but different flowgate usage, co-exist, leading to zero price on a congested flowgate, but these are not relevant to the analysis and would occur only rarely in practice.

237 This is always true for thermal constraints but not necessarily for stability constraints.

Thus, the LMP defined above is a *clearing price*: the generator is dispatched if its offer price is below the LMP and not dispatched if its offer price is above it.

Marginal Generators

Suppose that in an economic dispatch there is a *part-loaded* (ie, a generator which has some, but not all, of its offer dispatched) generator B at node B with offer price C_B. What happens to dispatch costs if another, zero-cost generator injects 1MW into node B. An obvious change to make is simply to reduce the output of generator B by 1MW. Since the total injection at node B – and hence the load flow - is the same as before, the dispatch must be feasible and the cost saving is C_B. Is this dispatch now *economic*, or is there a way of changing the dispatch so that the cost saving is more than C_B? If there were, that alternative dispatch would have been used originally, together with a 1MW increase in the output of generator B.

So, the cost saving in economic dispatch is C_B, meaning that the LMP is:

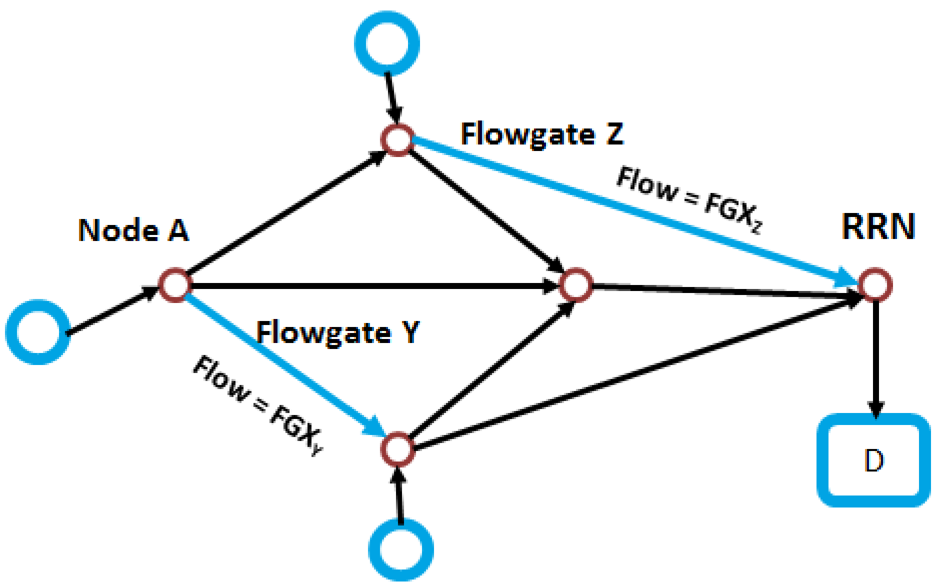
$$P_B = C_B$$

In general, whenever there is a part-loaded generator[238], the generator's offer price sets the LMP at its local node. Such a generator is referred to as *marginal*.

Relationship between Local and Flowgate Prices

Consider now the economic dispatch shown in Figure A.4, below. Flowgates Y and Z are both congested and have corresponding flowgate prices FGP_Y and FGP_Z.

Figure A.4 Economic Dispatch



238 Or strictly a generator that is part-loaded within a dispatch offer band.

What is the LMP at node A? The marginal value of generation at node A can be examined by considering changing economic dispatch in Figure A.4 above by adding a zero-cost generator producing 1MW at node A and a corresponding additional 1MW of demand at the RRN, with dispatch otherwise unchanged.

The superposition principle means that this is equivalent to *superimposing* on the original load flow a new load flow corresponding to 1MW from node A flowing to the RRN. The incremental output would flow through the two flowgates based on the node A participation factor. This would add to the flow already on the flowgates. Thus:

$$Flow_Y = FGX_Y + \alpha_{AY}$$

$$Flow_Z = FGX_Z + \alpha_{AZ}$$

Where:

α_{AY} is the participation of a generator at node A in flowgate Y

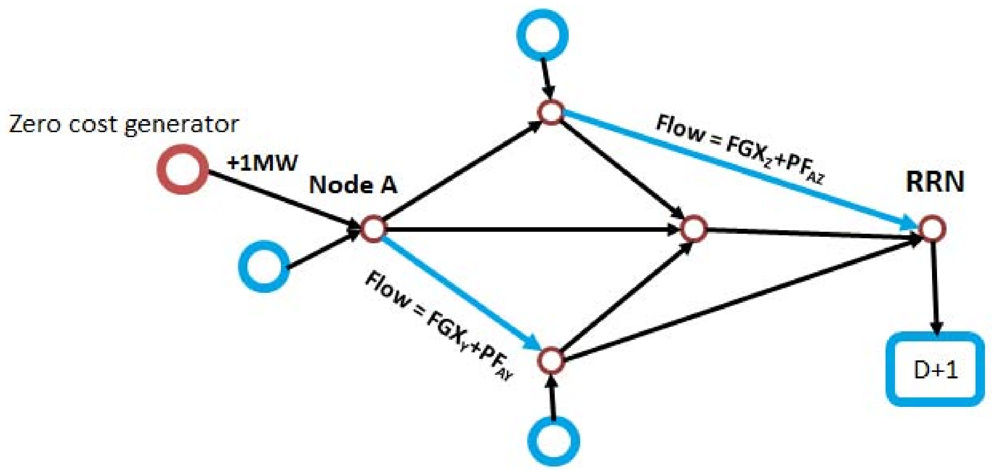
α_{AZ} is the participation of a generator at node A in flowgate Z

$Flow_Y$ is the flow through flowgate Y in the adjusted dispatch

$Flow_Z$ is the flow through flowgate Z in the adjusted dispatch

The adjusted dispatch and load flow is shown in Figure A.5.

Figure A.5 Adjusted Dispatch



This dispatch is *not feasible*, since the flowgate flows on Y and Z now exceed the flowgate capacity.[239] The dispatch must be changed: firstly, to reduce the flow on flowgate Y by α_{AY} and secondly to reduce the flow on flowgate Z by α_{AZ}. We know by

239 Assuming that the participation factors are positive. If they are negative, the dispatch is no longer economic, because the valuable flowgate capacity is being underutilised.

the definition of the flowgate prices that the cost of doing these two redispatches is $\alpha_{AY} \times FGP_Y$ and $\alpha_{AZ} \times FGP_Z$, respectively.

However, based on our definitions of LMP:

- the extra 1MW of generation at node A *decreases* dispatch costs by P_A; and
- the extra 1MW of demand at node R *increases* the dispatch cost by P_{RRN}.

P_{RRN} is the LMP at the regional reference node which, by definition, equals RRP.

Therefore, from the definition of flowgate prices:

$$\text{Net increase in dispatch costs} = \alpha_{AY} \times FGP_Y + \alpha_{AZ} \times FGP_Z$$

And from the definition of nodal prices:

$$\text{Net increase in dispatch costs} = RRP – P_A$$

So, putting the last two equations together:

$$RRP – P_A = \alpha_{AY} \times FGP_Y + \alpha_{AZ} \times FGP_Z \quad (A1.7)$$

Rearranging this equation, PA is defined by the formula:

$$P_A = RRP - \alpha_{AY} \times FGP_Y - \alpha_{AZ} \times FGP_Z \quad (A1.8)$$

The example considers the situation of two congested flowgates, but the analysis applies irrespective of the number of congested flowgates. In general, then, the LMP is defined by the formula:

$$LMP_i = RRP - \sum_k \alpha_{ik} \times FGP_k \qquad (A1.9)$$

Where:

LMP_i is the LMP at node i

α_{ik} is the participation of node i in flowgate k

FGP_k is the price of flowgate k

The summation can be over every *congested* flowgate or, equally, over *every* flowgate, recalling that the price of uncongested flowgates is always zero.

Inter-regional Prices

Suppose that, in the above example, the RRN was the reference node for region 2 and that node A was the reference node for region 1. Then equation (A1.7) becomes:

$$RRP_2 – RRP_1 = \alpha_Y \times FGP_Y + \alpha_Z \times FGP_Z \quad (A1.10)$$

The participation factors α_Y and α_Z represent the amount of flow through flowgates Y and Z for a flow from RRN_1 to RRN_2. Such a flow is referred to in the OFA model as the *directed interconnector* (DIC) from region 1 to region 2. Generalising equation (A1.10) gives:

$$RRP_N - RRP_S = \sum_k \alpha_k \times FGP_k \qquad (A1.11)$$

Where:

RRP_N is the RRP in the northerly region

RRP_S is the RRP in the southerly region

α_k is the participation of the northerly DIC in flowgate k

Local Price

The price paid to a generator, at the margin, for its output is referred to as the *local price.* That is to say, the total payment under regional and access settlement can be expressed in the form:

$$Pay\$ = fixed\$ + G \times P$$

Where:

Pay$ is the total settlement payment

G is generation output

Fixed$ is a dollar amount that is independent of G

P is the local price

Under OFA, a generator at node i is paid an amount:

$$Pay\$ = regional\ settlement\$ + access\ settlement\$$$

$$= RRP \times G + \sum_k\{(E_k\text{-}U_k) \times FGP_k\}$$

Where:

E = entitlement

U = usage

Suppose that the generator has non-negative participation in all congested flowgates: ie, it does not provide any flowgate support. Then the entitlement allocation process ensures that the entitlements, E, are all independent of generation. On the other hand, U is *not* independent of generation. Therefore, the fixed component of settlement consists of the payments relating to entitlements, ie:

$$\text{Fixed\$} = \sum_k (E_k \times FGP_k)$$

Equation (A1.12) can now be rewritten as:

$$\text{Pay\$} = \text{fixed\$} + G \text{ x } RRP - \sum_k (U_k \times FGP_k)$$

Usage is given by the formula:

$$U_k = \alpha_{ik} \times G$$

And so:

$$\text{Pay\$} = \text{fixed\$} + G \text{ x } \{(RRP - \sum_k (\alpha_{ik} \times FGP_k)\}$$

Using equation A1.9 to substitute the term in the curly brackets:

$$\text{Pay\$} = \text{fixed\$} + G \times LMP_i$$

So the local price, P, is equal to the LMP at node i. Recall that LMP is a clearing price, meaning that a generator is only dispatched if the LMP is equal to or higher than its offer price. Therefore, assuming that the offer price is equal to or higher than its short-run generating costs, the generator does not *regret* being dispatched.

Note that this result only applies to generators which do not provide any flowgate support on congested flowgates. This is because, for flowgate support generators, entitlements are *not* independent of dispatch.

A.1.4 Effective Access

Scaled Entitlements

In general, generators will not receive actual entitlements that are exactly equal to their target firm entitlements. When there are shortfalls, firm entitlements are scaled back; when there are surpluses, some non-firm entitlements may be allocated. Define scaling factors, s, to be the ratio of actual to target firm entitlements for a particular generator:

$$s_k = \text{actual entitlement on flowgate k / target firm entitlement on flowgate k}$$

$$= E_k / (\alpha_k \times A)$$

A flowgate support generator does not have a target entitlement but is instead allocated an entitlement equal to its usage:

$$E_k = \alpha_k \times G$$

Defining the scaling factor in the same way nevertheless gives:

$$s_k = E_k / (\alpha_k \times A) = G/A$$

To prevent a division by zero problem, it is assumed that A>0. It is seen from the above that s_k is non-negative.

Now, the total settlement payment for the generator is:

$$Pay\$ = RRP \times G + \sum_k\{(E_k\text{-}U_k) \times FGP_k\}$$

$$= RRP \times G + \sum_k\{(s_k \times \alpha_k \times A - \alpha_k \times G) \text{ x } FGP_k\}$$

$$= LMP \times G + A \times \sum_k(s_k \times \alpha_k \times FGP_k)$$

$$= LMP \times G + \{A \times \sum_k(s_k \times \alpha_k \text{ x } FGP_k)/(RRP\text{-}LMP)\} \times (RRP\text{-}LMP) \quad (A1.13)$$

$$= LMP \times G + A^e \times (RRP\text{-}LMP)$$

Where A^e is referred to as the *effective access*, and is given by the formula inside the curly brackets in equation (A1.13):

$$A^e = A \times \sum_k(s_k \times \alpha_k \times FGP_k)/(RRP\text{-}LMP)$$

$$= A \times \{\sum_k(s_k \times \alpha_k \times FGP_k)/ \sum_k(\alpha_k \times FGP_k)\}$$

$$= A \times s^e$$

Where s^e is the *effective scaling factor*, given by the formula:

$$s^e = \sum_k(s_k \times \alpha_k \times FGP_k)/ \sum_k(\alpha_k \times FGP^k)\} \quad (A1.14)$$

It is seen from equation A1.14 that s^e is the weighted average of the scaling factors s_i, with the weighting factor being the product of the participation factor and the FGP.

$$s^e = \sum_k(s_k \times w_k) / \sum_k w_k$$

Where:

w_k is the weighting factor on flowgate k

$$w_k = \alpha_k \times FGP_k$$

Examples of Effective Access

Equation A1.15 can be applied to some specific situations:

1. if there is congestion on only *one* flowgate in which the generator participates, the effective scaling factor equals the scaling factor on that congested flowgate;

2. if there are no *shortfalls* on congested flowgates, then each scaling factor will be equal to or greater than one, meaning that the weighted average must similarly be equal to or greater than one. Therefore, effective access is equal to or higher than registered access;

3. if there are no capacity surpluses on congested flowgates, the scaling factors for a non-firm generator will all be zero and so the generator has zero access;

4. where there is congestion on multiple flowgates, the effective scaling factor is weighted towards scaling factor on the flowgate with highest participation and flowgate price; and

5. a flowgate support generator has effective access equal to its output.

A.1.5 Mixed Constraints

Overview

Access settlement treats flowgates differently depending upon whether a generator or interconnector has a positive or negative participation in the flowgate. It is possible for a party to concurrently have negative and positive participation in different binding flowgates. The party is then said to be participating in *mixed constraints*.

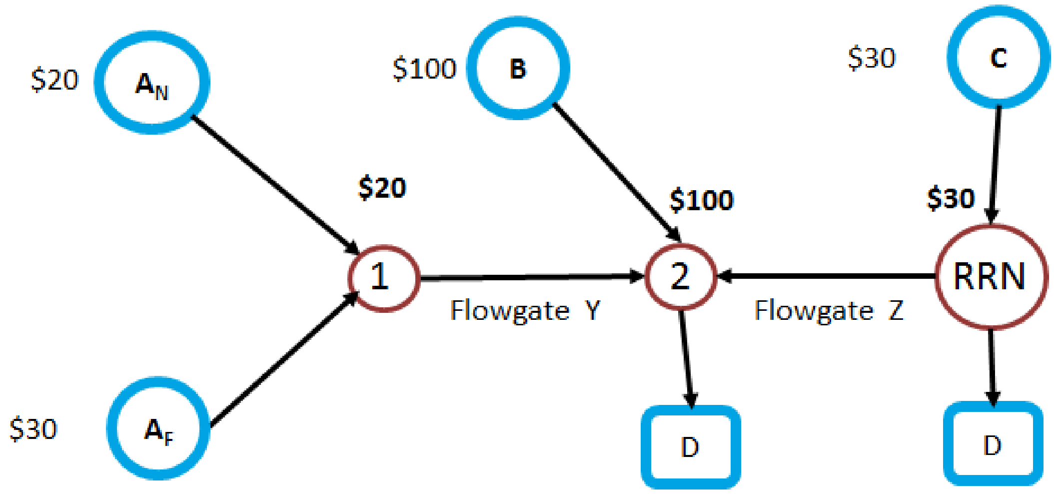
For generators, this scenario may be a largely theoretical situation which never, or rarely, occurs in practice. However, if quantitative modelling were to show that it could occur frequently, some changes may be needed to the OFA model design to address the issues arising, discussed in the next section. The situation is likely to be more common for an interconnector.

Generator Situation

Figure A.6, below, presents a simple scenario where generators participate in two binding flowgates.[240] Generators A_N and A_F, which are non-firm and firm, respectively, each have positive participation in flowgate Y and negative participation in flowgate Z, both of which are congested. Generator B is a simple flowgate support generator and C is a RRN generator.

240 This simple example may appear unrealistic because it relies on a generator, B, being constrained on. However, a similar mixed constraint combination could arise on a looped network without requiring constrained-on generation. The radial example is used because it is simpler to explain and understand.

Figure A.6 Mixed Generator Constraints



Assume that A_N receives zero entitlement on flowgate Y. It is therefore charged the flowgate price on its output.[241] However, because it is a flowgate support generator for flowgate Z, it has zero access settlement on this flowgate. Therefore, payments to generator A_N are:

Pay$$_N$ = RRN settlement$ + Flowgate Y settlement$ + Flowgate Z Settlement$

= $30 × G – ($100-$20) × G + 0$ = -$50 × G

So, in this example, a non-firm generator is not just paid less than its *offer price* for its output, it is paid less than *zero*. It is, in a sense, bearing the cost of being "constrained-on" (which all flowgate support generators bear), despite not actually being constrained on, given that its offer price is below the RRP.

The firm generator, A_F, is not dispatched (its offer is above the LMP), but receives an entitlement on flowgate Y. It is therefore paid FGPY ($80) on its access amount. It has zero access settlement on flowgate Z. Therefore, A_F is compensated at *$80* for being constrained off, despite there only being a $10 difference between the RRP and the LMP.

In summary, in the mixed constraint situation, neither a firm nor non-firm generator is paid in accordance with OFA principles:

- the non-firm generator (A_N) is not paid LMP;
- the firm, constrained-off generator (A_F) is not compensated in accordance with the difference between RRP and LMP; and
- the flowgate support generator (A_N) is not paid RRP.

241 Since the flowgate is radial, usage equals output.

Just as generators are rarely constrained-on *in practice* currently – they will rebid unavailable rather than be dispatched at a loss – rebidding is likely to reduce the instances of mixed generator constraints under OFA.

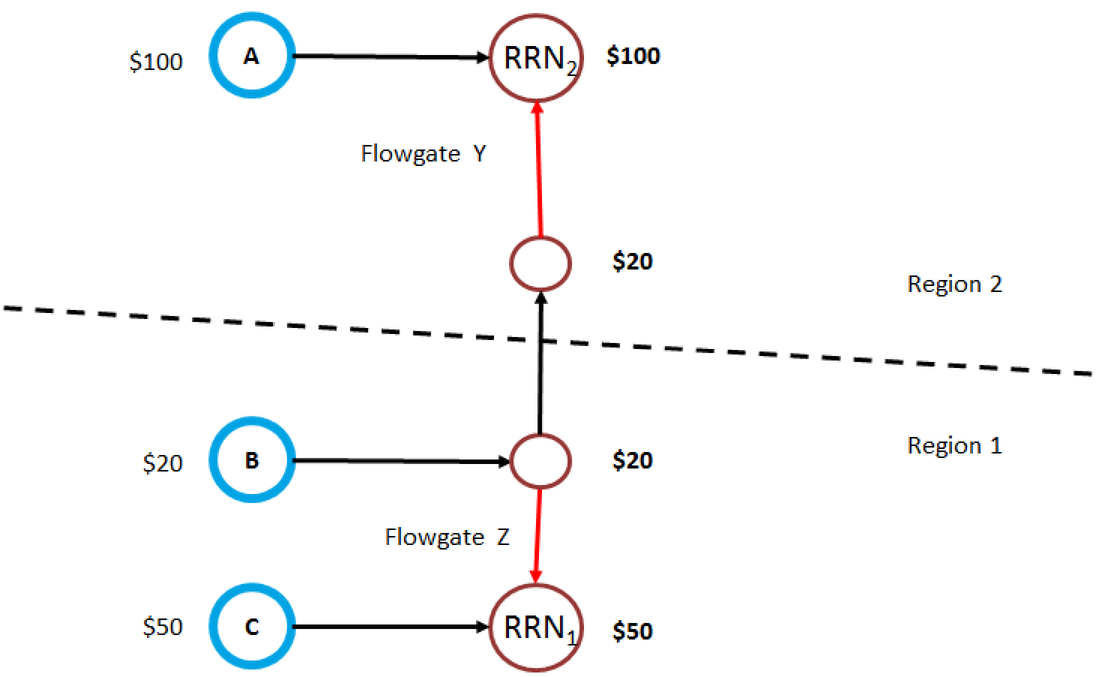
Generator A_N in this example would be likely to reduce its offered quantity until either flowgate Y became uncongested, which would mean the generator would be paid RRP, or its output reduced to zero. That should not materially affect system security: the flow on flowgate Y only needs to reduce by 1MW to remove the congestion. If, in a more complex example, generator A's backing off *did* create a security problem, AEMO could *direct* it and compensate it as necessary for any out-of-merit costs.

Once the congestion on flowgate Y is removed, the example reverts to a conventional flowgate support situation where generators A_N and A_F are paid RRP for their output.

Mixed Interconnector Constraints

Figure A.7, below, presents an example of an interconnector facing mixed flowgates.

Figure A.7 Mixed Interconnector Constraints



Recall that directed interconnectors usually have positive participation in flowgates, only having negative participation when there is negative flowgate capacity, which is not the case in the example. Therefore, in the example, the northerly interconnector participates in flowgate Y and the southerly interconnector participates in flowgate Z.

The addition of access settlement to existing IRSR provides a firm revenue stream on each flowgate. In the example, suppose that agreed inter-regional access is as follows:

- 100MW for the northerly interconnector; and

- 200MW for the southerly interconnector.

Assume also that firm entitlement targets are met and that there are no non-firm entitlements. In this situation, settlement payments to the DICs are as follows:

$$DIP\$_N = A_N \times FGP_Y = 100MW \times (\$100\text{-}\$20)$$

$$DIP\$_S = A_S \times FGP_Z = 200MW \times (\$50\text{-}\$20)$$

Where:

A is the agreed access

FGP is the flowgate price

DIP$ is the settlement payment to the DIC

suffixes N and S refer to northerly and southerly, respectively

suffixes Y and Z refer to flowgates Y and Z respectively

The settlement payouts to the northerly and southerly FIR holders are, therefore, based on price differences of $80 and $30, respectively, neither of which matches the inter-regional price difference of $50. That might suggest that the FIRs do not act as effective inter-regional hedges in this situation.

However, hedging is most critical when inter-regional price differences are extreme. Suppose that the price in the northerly region, RRP_2, increases to $10,000. Other things being equal, FGP_Y will increase to $9,980 which is (proportionately) very similar to the inter-regional price difference of $9,950. Therefore the northerly FIR acts as a very effective northerly inter-regional hedge under these severe conditions. On the other hand, if RRP_2 remains at $100 but now RRP_1 increases to $10,000, FGP_Z will increase to $9,980 – similar to the inter-regional price difference of $9,900 – so, similarly, the southerly FIR will be an effective southerly inter-regional hedge at such times.

Recall that, under the OFA model, DIC settlement does not depend upon the direction of interconnector *flow*.[242] In the situations described above the interconnector flow could be either northerly or southerly, depending upon the relative capacities of the two flowgates[243]; it would not substantially affect the DIC payments and so would not diminish the hedging effectiveness of the IRSRs. That compares to the status quo, where a change in interconnector flow direction (to a counterprice flow) would *completely remove* any hedging benefit from holding an SRA right.

242 The flow across the regional boundary.

243 If flowgates Y and Z were stability flowgates, or had capacity dependent on local demand or NSG, flowgate capacity could be quite variable.

A.2 DIC Settlement

A.2.1 Congestion Rent

Equation A1.11 defines the inter-regional price difference in terms of flowgate prices and is re-stated below:

$$RRP_N - RRP_S = \sum_k \alpha_k \times FGP_k$$

Ignoring the effect of losses, the IRSR is given by the formula:

$$IRSR\$ = I \times (RRP_N - RRP_S)$$

Where:

I is the northerly interconnector flow.

Combining these two equations gives:

$$IRSR\$ = I \times (\sum_k \alpha_k \times FGP_k)$$

$$= (\sum_k I \times \alpha_k \times FGP_k)$$

$$= (\sum_k U_k \times FGP_k) \quad (A1.16)$$

Where U_k is the interconnector usage of flowgate k. Recall that usage is the same whichever DIC is attributed to the flowgate, so U_k is also the DIC usage of the DIC attributed to flowgate k.

The individual components of the summation, $U_k \times FGP_k$, are the payments from the IRSR to the relevant DIC. Therefore:

IRSR\$ = total payment to DICs from IRSR

This ensures that settlement continues to balance: the IRSR is paid out in full to DICs.

When losses are included equation A1.16 does not hold, but is instead used to define two components of the IRSR:

Congestion rent\$ = $\sum_k (U_k \times FGP_k)$

Losses residue\$ = IRSR\$ - congestion rent\$

Providing the identity:

IRSR\$ = congestion rent\$ + losses residue\$

Therefore, once the DICs are paid out of the IRSR, the residual equals the losses residue.

A.2.2 FIR Settlement

DIC Payments

Payment to a DIC participating in a congested inter-regional flowgate is given by the formula:

$$DIP\$ = E \times FGP$$

Where:

E is the entitlement allocated to the DIC

This entitlement is made up of three components:

- a firm entitlement E^F;
- a non-firm entitlement E^N; and
- a (negative) flowgate support entitlement E^S: only allocated in the case of negative effective flowgate capacity.

Therefore:

$$DIP\$ = E^F \times FGP + E^N \times FGP + E^S \times FGP$$

= firm payment to FIR holders + non-firm payments to TNSP + flowgate support compensation paid by TNSP (A1.17)

The firm entitlement is calculated according to the formula:

$$E^F = s^F \times \alpha \times A^I \quad (A1.18)$$

Where:

s^F is the firm scaling factor (between zero and one)

A^I is the registered access of the DIC

α is the participation of the DIC in the flowgate

FIR Payments

The registered access of a DIC equals the aggregate registered FIR holdings on that DIC:

$$A^I = \sum_j Q_j^I \quad (A1.19)$$

Where:

Q_j^I is the quantity of the FIR on the DIC held by participant j

The FIR payment to each FIR holder on each congested is *defined* by the formula:

$$FIRpay\$_j = Q_j^I \times s^F \times \alpha \times FGP \quad (A1.20)$$

Where:

$FIRpay\$_j$ = payment to FIR holder j in relation to the inter-regional flowgate

Which means that the aggregate payment to FIR holders is:

$$TotalFIRpay\$ = \sum_j Q_j^I \times s^F \times \alpha \times FGP$$

$$= A^I \times s^F \times \alpha \times FGP \text{ (from equation A1.19)}$$

$$= E^F \times FGP \text{ (from equation A1.18)}$$

= settlement payment to DIC (from equation A1.17)

Therefore, using the formula in equation A1.20 to pay FIR holders ensures that settlement *balances*: that the payment to the DIC matches the aggregate payout to FIR holders. The balance occurs for each DIC on each congested flowgate and therefore DIC settlement balances globally.

Effective Inter-regional Access

The total payment, across all congested flowgates, to an FIR holder with holding Q is calculated by aggregating the individual payments defined in equation A1.20.

$$\text{Total FIR pay\$} = Q \times \sum_k s^F_k \times \alpha_k \times FGP_k \qquad (A1.21)$$

Where:

the suffix k indexes the variables for flowgate k

For simplicity, it is assumed that there is not mixed interconnector congestion and so the summation is over all congested flowgates in which the relevant interconnector participates.

It will be seen that this has a very similar form to equation A1.13, presented in relation to effective *intra*-regional access. Effective inter-regional access can be similarly defined:

$$Q^e = Q \times s^e \qquad (A1.22)$$

Where:

Q^e is the *effective inter-regional access* provided to the FIR holder

s^e is the *effective scaling factor*, defined by the equation:

$$s^e = \sum_k(s^F_k \times \alpha_k \times FGP_k) / \sum_k(\alpha_k \times FGP_k) \quad \text{(A1.23)}$$

Substituting s^e into equation A1.21 gives:

$$\text{Total FIR pay\$} = Q \times s^e \times \sum_k(\alpha_k \times FGP_k) \quad \text{(A1.24)}$$

Using equation A1.11, the summated expression in A1.24 is equal to the inter-regional price difference:

$$RRP_M - RRP_X = \sum_k \alpha_k \times FGP_k \quad \text{(A1.25)}$$

Where:

RRP_M is the RRP in the importing region (relative to the DIC direction)

RRP_X is the RRP in the exporting region (relative to the DIC direction)

Substituting from equations A1.25 and A1.22 into equation A1.24 gives:

$$\text{Total FIR pay\$} = Q^e \times (RRP_M - RRP_X)$$

So, the payout on the FIR equals the payout on an inter-regional hedge with quantity Q^e. Therefore effective access and effective scaling factor, presented here in the inter-regional context, have analogous meanings to those presented earlier in the intra-regional context.

In particular, s^e is a weighted average of the firm scaling factors applying on each congested inter-regional flowgate, with the weighting factor equalling the product of the participation and the flowgate price:

$$s^e = \sum_k(s^F_k \times w_k) / \sum_k w_k \quad \text{(A1.26)}$$

Where:

w_k is the weighting factor on flowgate k

$w_k = \alpha_k \times FGP_k$

Some general rules can be deduced from equation A1.26:

1. since the firm scaling factors are all between zero and one, the effective scaling factor must similarly be between zero and one: therefore the FIR payment is neither negative nor in excess of the *nominal* inter-regional hedge payment: $Q \times (RRP_M - RRP_X)$;

2. when the DIC participates in a single congested flowgate, the effective scaling factor is equal to the firm scaling factor on that flowgate; and

3. where the DIC participates in several congested flowgates, the effective scaling factor will be similar to the firm scaling factor on the flowgate with highest participation and flowgate price.

The firmness of inter-regional access – provided by an FIR – is determined by the same firm scaling factors as the firmness of intra-regional access. Therefore, under OFA, inter-regional and intra-regional access have equivalent firmness.[244]

A.3 Entitlement Allocation

A.3.1 EFGX cannot be negative

For the entitlement allocation process to be successful, effective flowgate capacity (EFGX) cannot be negative. EFGX is the sum of:

- the flowgate capacity;
- flowgate support from generators; and
- flowgate support from interconnectors.

On a congested flowgate, total usage equals flowgate capacity:

$$\sum_i U_i = FGX$$

The total usage can be divided into 3 components:

1. Total usage of flowgate access generators U^A: this is always non-negative
2. Usage of flowgate support generators U^S: this is always non-positive
3. Usage of interconnectors: U^I this may be positive or negative

Therefore:

$$U^A + U^S + U^I = FGX \quad (A1.27)$$

In the case where interconnectors are *not* required to provide flowgate support, EFGX is non-negative, by definition. Therefore, we only need to consider the case where both generators and interconnectors provide flowgate support. In this case the flowgate capacity is supplemented by the flowgate support:

$$EFGX = FGX + FS^G + FS^I$$

Where:

FS^G and FS^I represent the total flowgate support provided by generators and interconnectors respectively.

The flowgate support provided is equal to the negative of the usage:

244 They do not necessarily have *equal* firmness, since inter-regional flowgates may have a different actual level of firmness, for physical and geographical reasons, than intra-regional flowgates.

$FS^G = -U^S$

$FS^I = -U^I$

Substituting these into above definitions into equation A1.28 gives:

$EFGX = FGX - U^S - U^I$

Combining equations A1.27 and A1.29, we have:

$EFGX = U^A \geq 0$

Therefore, in this case also, EFGX is non-negative.

A.3.2 Target Access

In the core OFA design, firm and non-firm entitlements are calculated. In a possible modification to the core design, discussed in section 4.3.6, *super-firm* entitlements are calculated also. For the purposes of brevity, the calculation of super-firm entitlements is included in this appendix. To understand the calculation of entitlements in the core OFA design (ie, with no super-firm entitlements), all that is needed is to set the target super-firm access to zero.

Target access for generators is defined by the following formulae:

$A_F = \min(RA,GC)$

$A_{NF} = \max(GA-RA,0)$

$A_{SF} = \max(RA-GC,0)$

Where:

A is target access

RA = registered access

GA = generator availability

GC = generator registered capacity

F refers to the firm access component

NF refers to the non-firm access component

SF refers to the super-firm access component

DICs have no defined capacity or availability. Their target access is defined by the formulae:

$A_F = RA$

$A_{NF}=0$

$A_{SF} = 0$

Note that:

- all target access components are non-negative; and
- the sum of the firm and non-firm components for generators must equal or exceed availability.

The *target entitlement* for each component is calculated by multiplying the access component by the generator's participation factor for the relevant flowgate:

$ET_F = \alpha \times A_F$

$ET_{NF} = \alpha \times A_{NF}$

$ET_{SF} = \alpha \times A_{SF}$

A numerical example is provided in Table A.1 below to illustrate the target setting process. Note that each generator in the table belongs to a different access category. For simplicity, it is assumed that each generator is fully available.

Table A.1 Calculation of Target Entitlements

Generator	Nodal Values					α	Flowgate Values		
	RA	GC	A_F	A_{NF}	A_{SF}		ET_F	ET_{NF}	ET_{SF}
A (firm)	500	500	500	0	0	0.3	150	0	0
B (part-firm)	300	500	300	200	0	0.8	240	160	0
C (super-firm)	800	500	500	0	300	0.6	300	0	180
D (non-firm)	0	500	0	500	0	0.8	0	400	0
Total							690	560	180

The targets represent the *maximum* entitlements that generators will be allocated. In practice, one or more components will always be *scaled back*, through the entitlement scaling process described in the next section.

A.3.3 Actual Entitlements

For a flowgate to be congested, there must be the *potential* for total flowgate usage to be greater than flowgate capacity. Consider first an intra-regional flowgate. Recall that flowgate usage is:

$$U_i = \alpha_i \times G_i \qquad \text{(A1.30)}$$

Where:

U_i = flowgate usage of generator i

G_i = dispatch output of generator i

For there to be congestion, there must be some possible set of generator outputs, G_i, such that:

$$\sum_i (\alpha_i \times G_i) = \sum_i U_i > FGX \qquad \text{(A1.31)}$$

Now since, firstly, the sum of firm and non-firm target access exceeds availability and, secondly, availability equals or exceeds dispatched output:

$\sum_i ET_{Fi} + \sum_i ET_{NFi} \geq \sum (\alpha_i \times Avail_i)$ (since $A_F + A_{NF} \geq$ Avail as noted above)

$\geq \sum (\alpha_i \times G_i)$ (since $Avail_i \geq G_i$)

$> FGX$ (from equation (A1.31))

Therefore, if an intra-regional flowgate is congested, it is not possible to allocate all generators their firm and non-firm target entitlements and some scaling back is always required.

This is not necessarily true for an inter-regional flowgate. In this case, any remaining FGX after generators have been allocated their target entitlements in full is allocated to DICs in proportion to the product of their participation and their *capacity*.[245]

Entitlement scaling is based on the principles:

- total actual entitlements must equal flowgate capacity;
- a single *firm scaling factor* is applied to all firm and super-firm target entitlements, for generators and DICs, and a single *non-firm scaling factor* is applied to all non-firm target entitlements;
- firm entitlements are only scaled back when non-firm actual entitlements have been scaled back to zero;
- super-firm actual entitlements are only provided to the extent necessary to offset the scaling back of firm entitlements: ie, the sum of firm and super-firm actual entitlements is *never* higher than the firm target entitlement; and
- non-firm entitlements are only allocated to DICs when generators have been allocated their target non-firm entitlements in full.

The formulae for determining actual entitlements, based on target entitlements, are presented in Table A.2, below.

245 Capacity is defined for DICs in the same way as for generators, based on their maximum flow over the past two years.

Table A.2 Formulae for actual entitlements

Symbol	Meaning	Calculation
k_F	firm scaling factor	using a goal seek algorithm
k_{NF}	non-firm scaling factor	Min $\{1, k_{NF} = (FGX - \sum EA_F) / \sum ET_{NF}\}$
EA_F	actual firm entitlement (G and DIC)	$k_F \times ET_F$
EA_{NF}	actual non-firm entitlement (G only)	$k_{NF} \times ET_{NF}$
EA_{SF}	actual super-firm entitlement (G only)	$\min\{ET_F - EA_F, k_F \times ET_{SF}\}$
EA^I_{NF}	actual non-firm entitlement (DIC)	Sharing of any residual EFGX pro rata
EA	actual (total) entitlement	$EA_F + EA_{NF} + EA_{SF} + EA^I_{NF}$

To illustrate these formulae numerically, actual entitlements are calculated, from the targets presented in Table A.1, under two different scenarios:

- *scenario one*: low flowgate capacity; FGX = 522; and
- *scenario two*: high flowgate capacity; FGX = 802.

These outcomes are presented in Table A.3, below.

Table A.3 Actual entitlements under two capacity scenarios, with super-firm access

	Target Entitlements			Actual E: scenario 1 k_F=0.6; k_{NF}=0				Actual E: scenario 2 k_F=1; k_{NF}=0.2			
Generator	Firm	NF	SF	Firm	NF	SF	All	Firm	NF	SF	All
A	150	0	0	90	0	0	90	150	0	0	150
B	240	160	0	144	0	0	144	240	32	0	272
C	300	0	180	180	0	108	288	300	0	0	300
D	0	400	0	0	0	0	0	0	80	0	80
Total	690	560	180	414	0	180	522	690	112	0	802

In scenario 1, flowgate capacity (=522MW) is less than the aggregate firm target entitlements (=690MW). Therefore, since firm entitlements must be scaled back, no non-firm entitlements are provided. Note that generator C does not have its entitlements scaled back by as much as generator A does, because of the contribution from super-firm components.

In the above scenarios, super-firm entitlements are allocated. The situation in the core OFA design, with target super-firm access set to zero, is presented in Table A.4 below.

Table A.4 Actual entitlements under two capacity scenarios with no super-firm access (standard OFA design)

	Target Entitlements			Actual E: scenario 1 k_F=0.76; k_{NF}=0				Actual E: scenario 2 k_F=1; k_{NF}=0.2			
Generator	Firm	NF	SF	Firm	NF	SF	All	Firm	NF	SF	All
A	150	0	0	113	0	0	113	150	0	0	150
B	240	160	0	182	0	0	182	240	32	0	272
C	300	0	0	227	0	0	227	300	0	0	300
D	0	400	0	0	0	0	0	0	80	0	80
Total	690	560	0	522	0	0	522	690	112	0	802

A.4 Losses

A.4.1 Overview

Transmission losses are reflected in regional settlements through the application of marginal loss factors (MLFs) to dispatch and pricing. The treatment of losses in access settlement needs to align with regional settlement; in particular, to ensure that the principle of no-regret dispatch is maintained.

A.4.2 Marginal Loss Factors

In the current NEM design, *intra*-regional transmission losses are represented through static marginal loss factors (MLFs), which are defined for each node in a region. The MLF at a node represents the additional demand that can be supplied at the RRN with 1MW of incremental generation dispatched at that node.

$$\Delta D = (\text{Divergence})G \times MLF$$

Where:

ΔG is the incremental generation dispatched at a local node

ΔD is the incremental demand supplied at the RRN

MLF is the marginal loss factor applying to the local node

For example, if MLF=0.9, then 1MW of additional generation supplies only 0.9MW of extra demand at the RRN. The remaining 0.1MW is lost in the transmission network as a result of a marginal increase in transmission losses.

The value of demand at the RRN is, by definition, the RRP. Therefore, the value of the incremental generation, assuming no congestion, is:

$$\text{Incremental value} = \Delta D \times RRP = \Delta G \times MLF \times RRP$$

By definition, LMP is the marginal value of 1MW of generation at a node. Therefore, in this case:

$$LMP = \text{incremental value}/\Delta G = MLF \times RRP$$

In regional settlements, generator payments are loss-adjusted so that generators, in the absence of congestion, are paid this LMP.

$$\text{Regional Pay\$} = G \times MLF \times RRP$$

Where:

G is generation output

A.4.3 Dispatch Engine

Currently, NEMDE does not model losses explicitly but instead adjusts the offer prices that are submitted by generators, using the MLF:

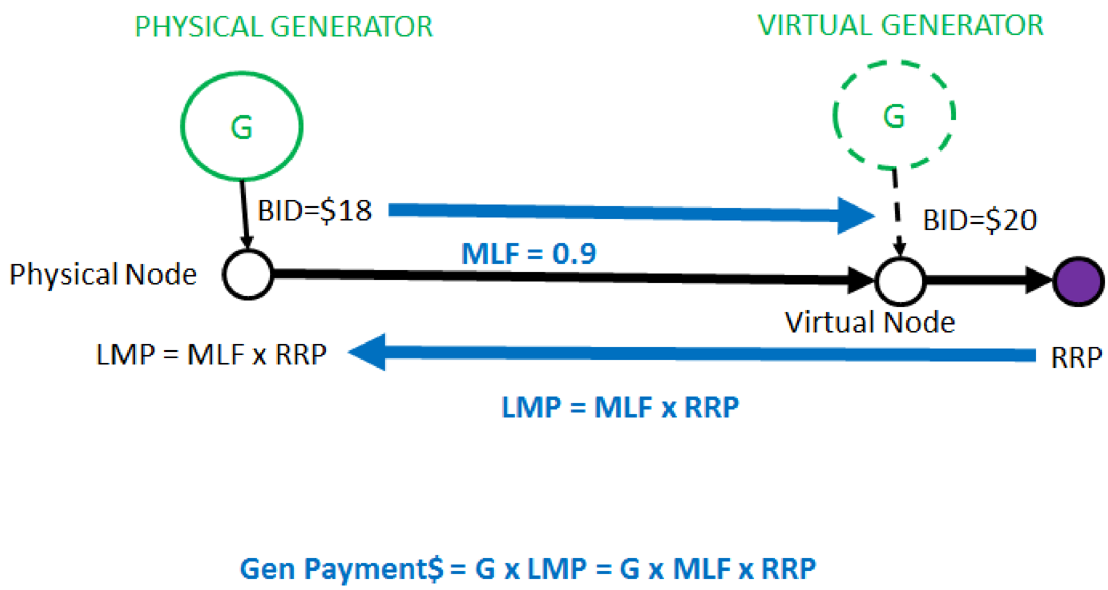
Adjusted offer price = offer price/MLF

It then dispatches each generator, in accordance with the adjusted offer price.

To illustrate this, consider a generator with an offer price of $18 and an MLF of 0.9, giving it an adjusted offer price of $20. In a sense, NEMDE does not dispatch the generator with losses, but rather an equivalent virtual generator that is located at a virtual node close to the RRN, from which there are no losses. This is illustrated in Figure A.8.

The generator will be dispatched if the RRP is $20 or higher, meaning that MLF x RRP is $18 or higher. Therefore, these arrangements comply with no-regret dispatch, in which a generator is dispatched only when it is paid higher than its offer price, at the margin.

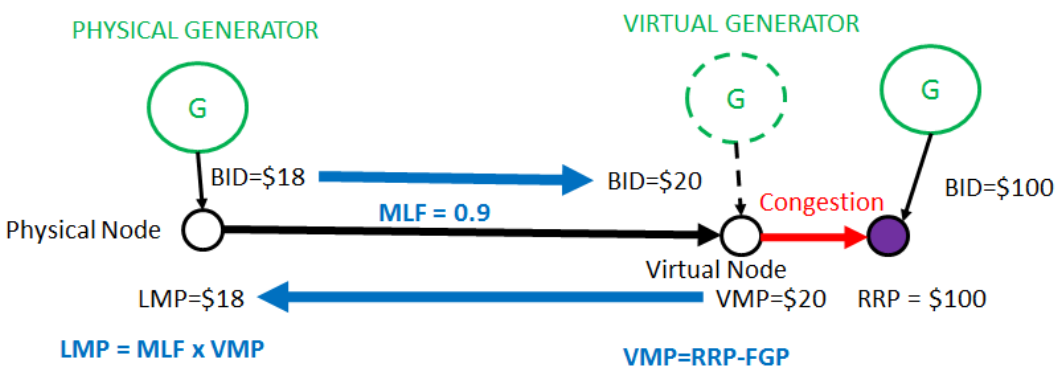
Figure A.8 Loss Factors in Uncongested Dispatch



A.4.4 Flowgate Prices

This situation changes currently when there is intra-regional congestion. Because it models a lossless network, NEMDE implicitly assumes that congestion occurs between the virtual node and the RRN, as illustrated in Figure A.9. In the example, congestion causes the local generator to be constrained off. RRP is maintained at $100 due to a marginal generator located at the RRN. As before, the local generator has an offer price of $18 and an MLF of 0.9. This is converted to a virtual generator located at the virtual, lossless node with an offer price of $20.

Figure A.9 Loss Factors under Congested Dispatch



If the capacity on the congested flowgate were increased by 1MW then:

- the flow through the flowgate could increase by 1MW, offsetting 1MW of the more expensive RRN generation; and
- to supply this 1MW, the virtual generator has to increase its output by 1MW.

Therefore, the flowgate price is the consequent reduction in dispatch cost is:

$$FGP = \text{Dispatch saving} = \$100 \times 1MW - \$20 \times 1MW = \$100 - \$20 = \$80$$

The virtual marginal price (VMP), the LMP at the virtual node, is the offer price of the virtual generator. Therefore:

$$VMP = RRP - FGP$$

The LMP is related to the VMP according to the formula:

$$LMP = MLF \text{ x } VMP$$

Therefore:

$$LMP = MLF \times (RRP - FGP)$$

The generalised formula for LMP in a lossless network was derived in A.1. This formula now applies to VMP:

$$VMP_i = RRP - \sum_k \alpha_{ik} \times FGP_k$$

The LMP is just the VMP adjusted for losses, and so:

$$LMP_i = MLF \times (RRP - \sum_k \alpha_{ik} \times FGP_k)$$

A.4.5 Usage Adjustment

In a lossless network, access settlement on a congested flowgate under OFA would use the formula:

$$Pay\$ = (E\text{-}U) \times FGP$$

Where:

$$U = \alpha \times G$$

When there are losses, the formula applies to the *virtual* generator:

$$U = \alpha \times G_{virtual}$$

The virtual generation is related to the actual generation by the MLF:

$$G_{virtual} = MLF \times G$$

Therefore, the formula for U is:

$$U = \alpha \times G_{virtual} = \alpha \times MLF \times G$$

Therefore, usage must be loss adjusted.

With this adjustment, the no-regret dispatch principle still applies for flowgate access generators, as demonstrated below.

The total settlement payment to a generator is:

$$Pay\$ = regional\$ + access\$$$

$$= MLF \times RRP \times G + \sum_k FGP_k \times (E - U)$$

$$= MLF \times RRP \times G + \sum_k FGP_k \times (E - \alpha_k \times MLF \times G)$$

E is set independent of dispatch, so the E-related payment can be expressed as a fixed payment and the other terms can be re-arranged, giving:

$$Pay\$ = fixed\$ + G \times MLF \times \{RRP - \sum_k (\alpha_k \times FGP_k)\}$$

$$= fixed\$ + G \times LMP$$

Therefore, the local price paid to the generator at the margin equals the loss-adjusted LMP.

A.4.6 Clearing Price

As noted previously, NEMDE does not model losses but instead dispatches against loss-adjusted offer prices. Therefore, there is a lossless dispatch in which:

- the generator offer price is loss-adjusted: offer price/MLF;
- a lossless VMP is calculated at the virtual node; and

- the lossless VMP is a clearing price: the generator is dispatched if its loss-adjusted offer price is less than the VMP.

Therefore, if a generator is dispatched:

$$\text{VMP} \geq \text{adjusted offer price} = \text{Offer price}/\text{MLF}$$

And so:

$$\text{LMP} = \text{MLF} \times \text{VMP} \geq \text{Offer price}$$

In conclusion:

- LMP is a clearing price;
- the generator is paid LMP at the margin; and
- therefore, the no-regret dispatch principle is satisfied.

B Access Settlement Practicalities

B.1 Generator Configurations

B.1.1 Terminology

This section uses terminology that is defined below.

A *dispatchable unit* is the representation of generation in existing dispatch processes. Generally, a dispatchable unit is simply a physical generating unit. However, in some cases, a dispatchable unit is an *aggregated unit*: ie, a notional aggregation of multiple generating units. Dispatchable units are labelled as “DUID” (dispatchable unit identifier) in the figures.

A *SCADA meter* is a meter that is used in dispatch but not in regional settlement. Although they are shown in the illustrations, SCADA meters are not used in access settlement and so can be ignored for the purposes of this section.

A *revenue meter* is a meter that is used in regional settlement and will similarly be used in access settlement. When a meter is referred to in this section, therefore, it is implied that it is a revenue meter.

As-generated output means the gross output of a generating unit.

Auxiliary load is load that is used in power station operation and which, in access settlement, is deducted from as-generated output in order to calculate a *sent-out* measure of generation output.

B.1.2 Access Settlement Principles

The following principles are used in the OFA design:

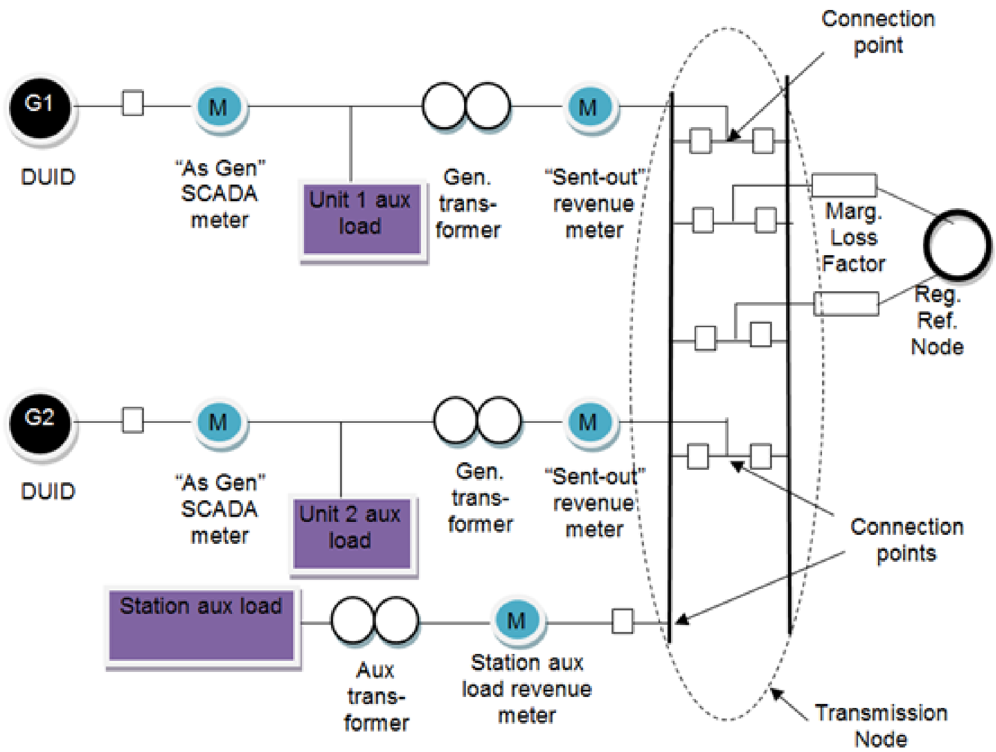
1. settlement is on a sent-out basis: with *auxiliary load* being deducted from the *as-generated* output;
2. settlement is by *access unit*: an *access unit* is usually the same as a dispatchable unit, but consists of multiple dispatchable units where these share metering; and
3. only revenue metering is used: ie, high accuracy metering, currently used for regional settlement, which measures half-hourly quantities.

These principles have been introduced and explained elsewhere in chapter 4. This section illustrates how these principles are *applied* in the context of various generation connection and metering configurations.

B.1.3 Standard Configuration

A standard connection configuration is presented in Figure B.1.

Figure B.1 Standard Configuration



In this example:

- there are two generating units at a power station;
- there are two dispatchable units, corresponding to the two generating units;
- each dispatchable unit has a separate *unit meter*;
- the unit meter measures as-generated minus unit load; and
- station load is metered by a separate *auxiliary meter*.

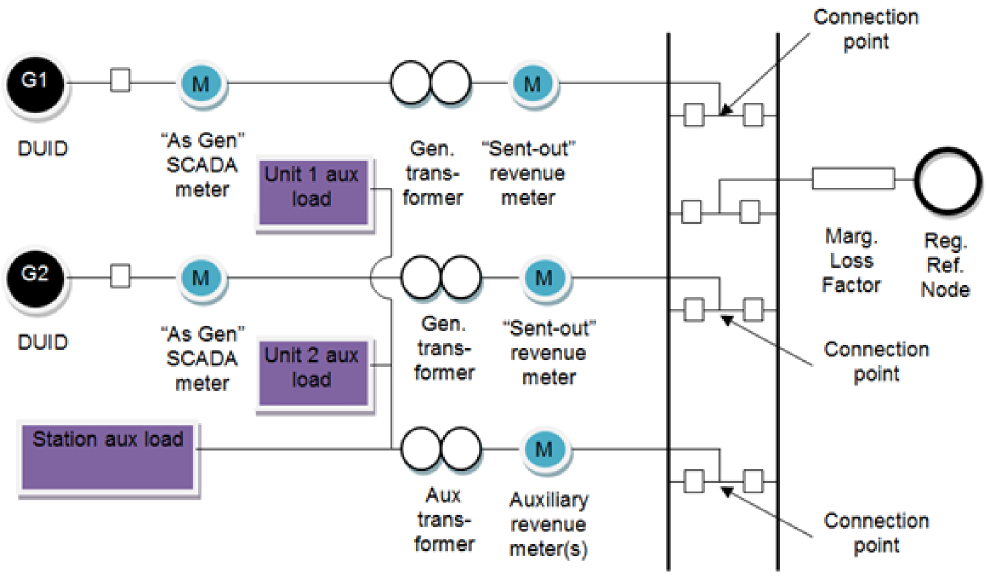
The approach to access settlement in this example is as follows:

- there are two access units, corresponding to the two dispatchable units;
- the station load must be allocated between the two access units: eg, in proportion to the unit metered quantities; and
- sent-out generation for an access unit equals the unit metered quantity minus the share of the station load.

B.1.4 Combined Metering of Auxiliary Load

A combined metering connection configuration is presented in Figure B.2.

Figure B.2 Combined Metering of Auxiliary Load



In this example:

- there are two generating units at a power station;
- there are two dispatchable units, corresponding to the two generating units;
- each dispatchable unit has a separate *unit meter*;
- the unit meter measures as-generated; and
- combined unit and station load is metered by a separate auxiliary meter.

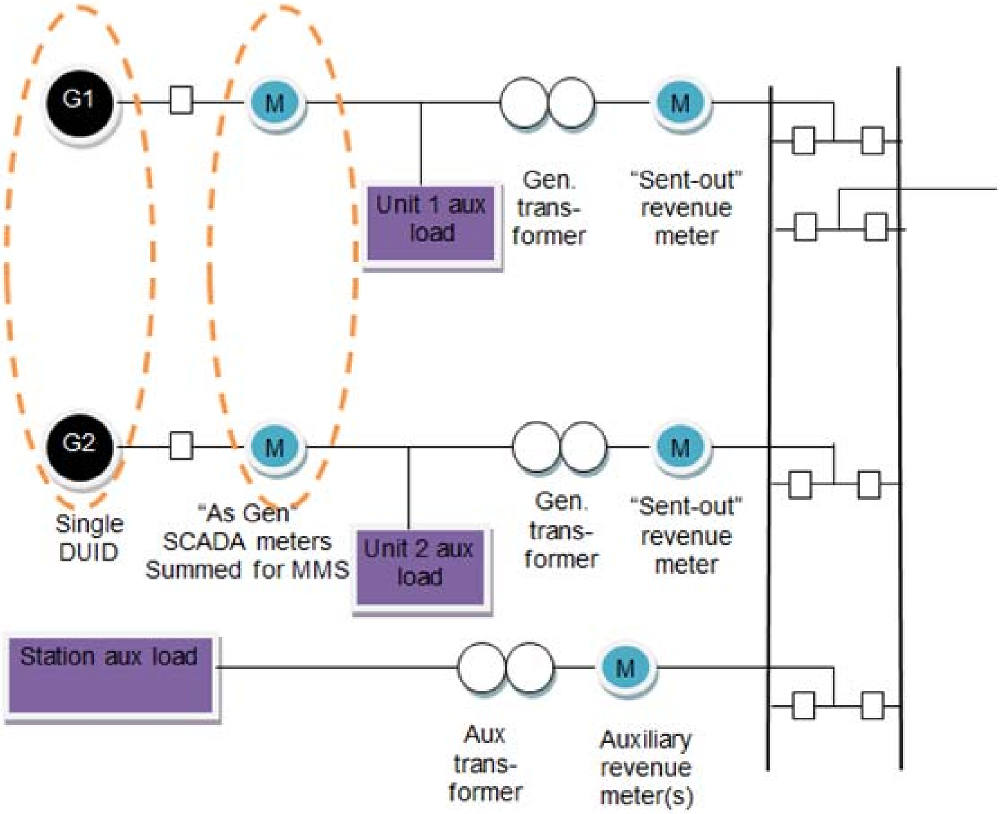
The approach to access settlement in this example is as follows:

- each dispatchable unit is a separate access unit;
- the auxiliary load must be allocated between the two access units: eg, in proportion to the unit metered quantities; and
- sent-out generation for an access unit equals the unit metered quantity minus the share of the auxiliary load.

B.1.5 Aggregated Units

A connection configuration involving aggregated units is presented in Figure B.3.

Figure B.3 Aggregated Units



In this example:

- there are two generating units at a power station;
- there is a single, aggregated dispatchable unit, being the aggregate of the two generating units;
- each generating unit has a separate unit meter;
- the unit meter measures as-generated minus unit load; and
- station load is metered by a separate auxiliary meter.

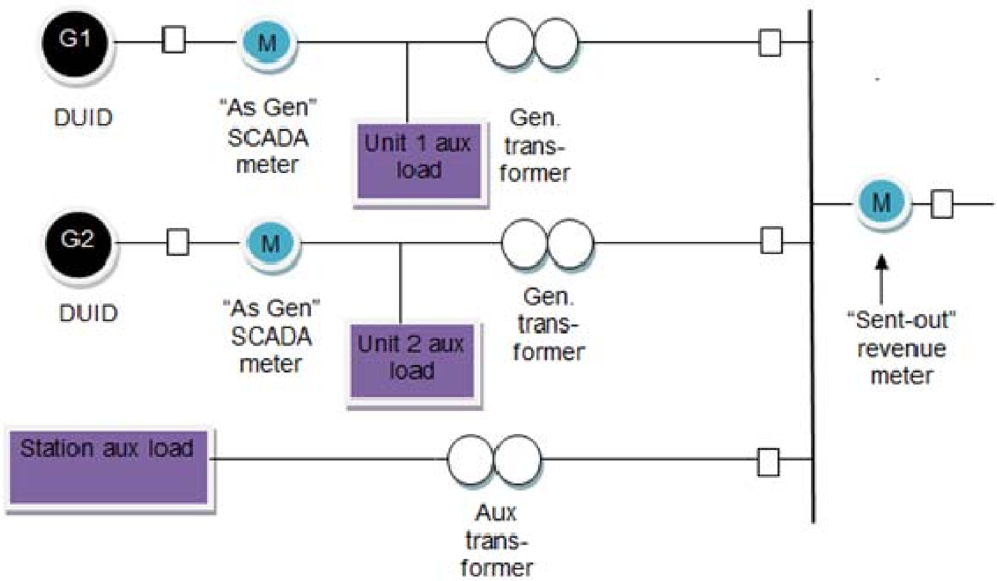
The approach to access settlement in this example is as follows:

- there is a single access unit, corresponding to the single, aggregated dispatchable unit; and
- sent-out generation for the access unit equals the aggregate of unit metered quantities minus the auxiliary metered quantity.

B.1.6 Combined Output Metering Unit

A connection configuration involving combined output metering is presented in Figure B.4.

Figure B.4 Combined Output Metering



In this example:

- there are two generating units at a power station;
- there are two dispatchable units, corresponding to the two generating units; and
- there is a single meter which measures combined as-generated output minus combined unit and station loads.

The approach to access settlement in this example is as follows:

- there is a single access unit, which is an aggregation of the two dispatchable units; and
- sent-out generation for the access unit equals the metered quantity.

B.2 Flowgates and Dispatch Constraints

B.2.1 Overview

Flowgates correspond to transmission constraints used by NEMDE. A transmission constraints is defined as:

> "any constraint that arises as a result of limitations on TNSP networks, or DNSP networks to the extent they involve dual function assets, and for which a constrained generator is not compensated under current arrangements."

This section considers the various categories of NEMDE constraint and assesses whether these are transmission constraints.

AEMO lists three broad categories of constraint:[246]

- network;
- frequency standards; and
- other.

These are considered in turn below.

B.2.2 Network Constraints

Network constraints cover three constraint types: thermal constraints, stability constraints[247] and network control schemes.[248] The former two types would certainly be treated as flowgates. The third type would probably not be treated as such, as generators are currently paid for providing Network Control Ancillary Services (NCAS).

B.2.3 FCAS constraints

FCAS constraints (those relating to frequency standards) are not generally caused by limitations on TNSP networks, meaning that they are not considered flowgates. Many FCAS constraints are affected by Basslink limitations, but Basslink is a MNSP rather than a TNSP and so, again, these do not give rise to flowgates in access settlement.

One type of FCAS constraint that is relevant to the OFA model is the *separation constraint*. Such a constraint may be included in NEMDE in situations where a credible contingency can lead to islanding.[249] A separation constraint sets a limit on the pre-contingent flow on the relevant network element to ensure that, should it fail, the FCAS in the two post-continent islands can contain frequency deviations in accordance with NEM operating standards.

For the purposes of OFA, a separation constraint is similar to a thermal network constraint that limits the pre-contingent element flow to a specified maximum. The difference in NEMDE is that the separation constraint is *co-optimised*, meaning that NEMDE can decide to source extra FCAS in order to increase the flow limit. This co-optimisation is not relevant to access settlement, which takes the flow limit (and associated flowgate capacity) at face value[250] and applies the access settlement algebra accordingly.

---

246 AEMO, Constraint Formulation Guidelines, 6 July 2010.

247 Covering voltage stability, transient stability and oscillatory stability.

248 Described by AEMO as the modelling of generator control schemes or reactive control devices on generator output.

249 The splitting of the NEM into two or more separated networks.

250 Recalling that flowgate capacity on binding constraints is calculated based on aggregate usage.

In summary, separation constraints are treated as transmission constraints in OFA, but other FCAS constraints are not.

B.2.4 Other constraints

AEMO lists the following *other* types of constraints:

- managing negative residues (during interconnector counterprice flows);
- rate of change (of interconnector or generator output);
- non-conformance;
- network support agreement;
- unit zero constraint (a generator is unable to generate – eg, due to transmission limitations - but is not bid as unavailable); and
- discretionary limit on generators or interconnectors.

These are very specific and technical constraints and decisions on whether to treat these as flowgates may sometimes need to be taken on a case-by-case basis. However, applying the informal definition above would suggest that:

- any constraints on *regulated interconnector* flows are transmission constraints: ie, those relating to managing negative residues,[251] interconnector rate of change limits and discretionary limits;
- unit zero constraints might be considered to be transmission constraints, where they relate to network limitations;
- network support agreements generally impact only on flowgate support generators and so would not need to be treated as flowgates in access settlement;[252]
- constraints relating solely to generator limitations or non-conformance are not transmission constraints; and
- discretionary limits on generators may be treated as transmission constraints where they arise as a result of network limitations.

B.2.5 Constraint Tagging

NEMDE constraints would need to be *tagged* as either "transmission constraint" or "not transmission constraint" in order that access settlement processes extract and

251 Although these are unlikely still to be required under an OFA regime.

252 Flowgate support generators receive zero payments from access settlement, so settling flowgates in which only support generators participate is unnecessary.

process only the relevant constraint information. Logically, this should be done by AEMO, who prepares the constraints.

B.3 Settlement Period

B.3.1 Terminology and Assumptions

In relation to the settlement period, three alternative access settlement mechanisms are discussed:

1. *(unweighted) TI*: settlement on a trading interval (TI) basis, based on unweighted averages of dispatch variables;

2. *weighted TI*: settlement on a trading interval basis, based on weighted averages of dispatch variables: the FGP is used as the weighting factor; and

3. *DI:* settlement on a *dispatch interval* basis.

The core design proposes the use of unweighted TI. However, all three options are considered in this section.

A trading interval is a half-hour. A dispatch interval is five minutes. So, there are six DIs making up each TI.

The access settlement algebra is presented below, using the usual terminology in Table B.1.

Table B.1 Terminology for Access Settlement Algebra

Symbol	Meaning
G	Sent out generation
U	Usage
FGP	Flowgate price
EFGX	Effective flowgate capacity
TFGX	Target flowgate capacity
A	Registered access
Pay$	Access settlement payment

Variables with a subscript, i, are *dispatch interval* variables, with i representing the DI within the TI. Variables *without* a subscript are trading interval variables. Where a TI variable has a superscript "w" it is a *FGP-weighted* average of the DI values. If it has no superscript then it is an *unweighted* average.

TI Payment variables have a DI, TI or WTI superscript, according to the settlement method used.

Loss factors are ignored. It is assumed that no generators are super-firm (A>capacity), so capacity can also be ignored.

For convenience, the quantity variables (G, U, E etc) are in units of MW, rather than MWh. So, the TI settlement is based on the formula:

$$Pay\$ = (E-U) \times FGP/2$$

DI settlement is based on the formula:

$$Pay\$_i = (E_i-U_i) \times FGP_i/12$$

$$Pay\$ = \sum_i Pay\$_i$$

The following simplifying assumptions are made. These are examined later:

1. $EFGX_i < TFGX$, in every DI, so no non-firm entitlements are allocated;

2. participation factors are fixed over a trading interval; and

3. $EFGX_i$ does not change sign within the trading interval.

It is also assumed that registered access is constant over a trading interval. That is inevitable: firm access terms will be a complete TI, *at the very least*, and most likely *much* longer.

Recall that TFGX is the aggregate of target firm entitlements, based on the formula:

$$E = \alpha \times A$$

Therefore, since α and A are assumed not to change within the TI, neither does TFGX.

B.3.2 DI settlement

In DI settlement, values of G for each DI need to be *inferred* from the (revenue) meter reading, which is a TI value. The *profile* of the G_i values could be based on either SCADA metering or dispatch targets. In each case, the G_i values are scaled so that:

$$\sum_i G_i = 6 \times G$$

Entitlements are allocated for each DI, based on $EFGX_i$. The target entitlements need to be scaled back, based on the shortfall ratio: $EFGX_i/TFGX$. Therefore:

$$E_i = EFGX_i\ /TFGX \text{ x target firm entitlement} = EFGX_i\ /TFGX \times \alpha \times A$$

Usage is based on participation and DI generation:

$$U_i = \alpha \times G_i$$

Therefore, for DI settlement:

$$Pay\$_i = (E_i - U_i) \times FGP_i/12$$

$$= \alpha \times A / TFGX \times EFGX_i \times FGP_i/12 - \alpha \times G_i \times FGP_i / 12$$

$$Pay\$^{DI} = \sum_i Pay\$_i$$

$$= \alpha \times A / TFGX \times \sum_i (EFGX_i \times FGP_i) / 12 - \alpha \times \sum (G_i \times FGP_i) / 12$$

Now:

$$\sum_i (EFGX_i \times FGP_i) / 12 = \sum_i FGP_i / 12 \times \sum [EFGX_i \times FGP_i] / \sum_i FGP_i$$

$$= FGP/2 \times EFGX^w$$

And similarly:

$$\sum (G_i \times FGP_i/12) = G^w \times FGP/2$$

Therefore:

$$Pay\$^{DI} = \{A \text{ x } EFGX^w / TFGX - G^w\} \times \alpha \times FGP/2 \quad (A2.1)$$

B.3.3 TI settlement

In unweighted-TI settlement, entitlements are calculated for each TI, based on EFGX. Again, firm entitlements are scaled back in proportion to the shortfall:

$$E = EFGX / TFGX \times \alpha \times A \qquad (A2.2)$$

EFGX is an unweighted average of $EFGX_i$:

$$Pay\$^{TI} = (EFGX / TFGX \times \alpha \times A - \alpha \times G) \times FGP / 2$$

For weighted-TI settlement, the variables U and EFGX are based on weighted-averages of the DI-sourced variables; FGP is based on the unweighted average. The entitlement is calculated using equation A2.2 above, but using the weighted-average EFGX instead:

$$E = EFGX^w/TFGX \times \alpha \times A$$

Now, because α is assumed to be constant over the TI:

$$U^w = \alpha \times G^w$$

Therefore:

$$Pay\$^{WTI} = (E - U^w) \times FGP/2$$

$$= (EFGX^w / TFGX \times \alpha \times A - \alpha \times G^w) \times FGP / 2$$

This is identical to the formula (equation A2.1) that was derived for DI settlement. Therefore, under the assumptions listed above, weighted-TI settlement gives identical outcomes to DI settlement:

$$Pay\$^{DI}=Pay\$^{WTI}$$

B.3.4 Non-firm Entitlements

The simplifying assumptions used above mean that no non-firm entitlements are allocated. This means that the same level of (firm) entitlements were allocated under weighted-TI and DI settlement.

Suppose instead that $EFGX_i$ exceeded TFGX in some DIs but, nevertheless, $EFGX^W$ was less than TFGX. In this case:

- some non-firm entitlements will be allocated in DI settlement; and
- no non-firm entitlements will be allocated in WTI settlement.

Now, the *value* of an entitlement in settlements is E x FGP. Since entitlements must sum to EFGX, the total value of entitlements in DI settlement is:

$$Evalue^{DI} = \sum_i Evalue_i = \sum_i EFGX_i \times FGP_i \; / \; 12 = EFGX^w \times FGP \; / \; 2$$

This equals the total value of entitlements in weighted-TI settlement and so:

$$Evalue^{DI}= Evalue^{WTI}$$

In summary:

- the total value of entitlements is the same in DI and weighted-TI settlement;
- non-firm entitlements are liable to be higher in DI settlement; and
- by implication, the value of firm entitlements is likely to be lower in DI settlements.

It seems *inappropriate* that non-firm entitlements should be provided when, within the same TI, firm entitlements are scaled back. In this respect, DI settlement appears to be *inferior* to weighted-TI settlement.

There are analogous implications if $EFGX_i$ changes sign within a TI. Suppose that $EFGX_i<0$ in some DIs but, nevertheless, $EFGX^W$ is positive. Under TI settlement, DICs will be treated conventionally. However, under DI settlement – for those DIs with a negative $EFGX_i$ – DICs will provide flowgate support, with the associated cost charged to TNSPs. Firm entitlements will be *higher*, overall, as a result. In this case, it is not clear which outcome is preferable.

B.3.5 Unweighted TI vs Weighted TI

Recall the formulae for weighted and unweighted TI settlement:

$$Pay\$^{TI} = \alpha \times \{A \times EFGX / TFGX \text{ x } FGP / 2 - G \times FGP / 2\}$$

And:

$$Pay\$^{WTI} = \alpha \times \{A \times EFGX^{w} / TFGX \times FGP / 2 - G^{w} \times FGP / 2\}$$

Clearly, the payments are different only to the extent that $EFGX^{w}$ and G^{w} differ from EFGX and G, respectively: ie, the weighted averages differ from the unweighted averages.

Any differences will arise from variations in G_i and $EFGX_i$ over the trading interval. The sign of the difference depends upon the *correlation* between these variations and the variations in FGP:

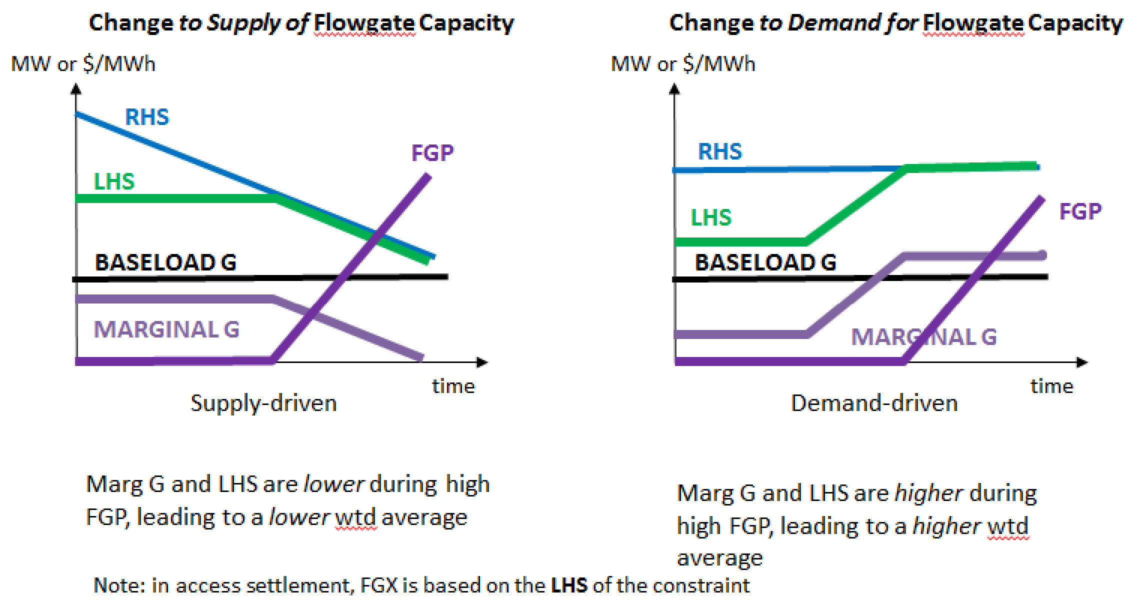
- if FGP variations are *positively* correlated with variations in G or EFGX, then the weighted averages will be *higher* than the unweighted averages; and
- if FGP variations are *negatively* correlated with variations in G or EFGX, then the weighted averages will be *lower* than the unweighted averages.

Changes in FGP could be *supply driven* or *demand driven*. Consider situations where a flowgate becomes congested midway through a TI. Recall that, in access settlement, EFGX is based on total usage of flowgate access generators: ie, the LHS of the constraint equation. When a flowgate is uncongested in a DI, LHS<RHS and so the estimated EFGX will be lower than the true EFGX. Supply-driven and demand-driven outcomes are illustrated in Figure B.5, which is explained below.

In a *supply-driven* change (presented in the left-hand graph in Figure B.5), a fall in EFGX creates congestion and so FGP rises. Marginal constrained generators will be backed off in order to manage the congestion. Therefore, both EFGX and G (for marginal generators) are *negatively* correlated with FGP. Baseload G does not change its output. The "LHS" and "RHS" curves show the left-hand side and right-hand side of the relevant constraint equation, respectively. The RHS is permitted to exceed the LHS, but not vice versa.[253] For simplicity, flowgates support is ignored, so FGX=EFGX.

253 It is implicitly assumed that the relevant constraint is invoked throughout the TI.

Figure B.5 Supply-Driven Vs. Demand-Driven Changes



On the other hand, in a *demand-driven* change, marginal generators rebid in an attempt to increase their output. This causes a previously uncongested flowgate to become congested. Thus, both generator output and EFGX rise, as FGP rises; they are *positively* correlated with FGP. This is illustrated in the right-hand graph in Figure B.5.

The implications for the relative values of the weighted and unweighted averages for EFGX and G are summarised in Table B.2, below. Note that because a baseload generator does not change its output over the TI, its weighted and unweighted output averages are always the same.

Table B.2 Weighted and Unweighted Averages

	Supply-driven changes	Demand-driven changes
Flowgate capacity	$EFGX^W<EFGX$	$EFGX^W>EFGX$
Marginal generator	$G^W<G$	$G^W>G$
Baseload generator	$G^W=G$	$G^W=G$

Firm generators gain from higher EFGX through getting the value of the higher entitlements. Non-firm generators are indifferent to EFGX, assuming that no non-firm entitlements are allocated. All generators gain from a lower G, since this leads to lower usage. In the light of these facts, the impacts for various generator types of using WTI settlement, rather than TI settlement are summarised in the table below. Where there is no impact, the cell is left blank.

Table B.3 G and EFGX Impacts

G type	Supply Driven		Demand Driven	
	G impact	EFGX impact	G impact	EFGX impact
Baseload non-firm				
Marginal non-firm	Gain		Lose	
Baseload firm		Lose		Gain
Marginal firm	Gain	Lose	Lose	Gain

The relative frequency and materiality of supply-driven and demand-driven congestion is unknown. The relativity is important, however, since it will determine the *net* impact of the choice of settlement method on the various generator types.

One interesting outcome is that, under weighted-TI settlement, a marginal non-firm generator loses under demand-driven congestion. Thus, using weighted-TI settlement may deter somewhat any rebidding that is *designed* to cause congestion.

As noted above, under the simplifying assumptions, DI and weighted-TI settlement give identical outcomes. Therefore, the impact of using DI settlement is likely to be similar to that of using weighted-TI settlement.

B.3.6 Incentive Scheme Settlement

The operational incentive scheme places penalties or rewards on TNSPs based on the level and cost of shortfalls: the amount, if any, by which EFGX falls short of TFGX. This raises the question of whether a weighted or unweighted value of EFGX should be used.

A key difference between the two approaches is that, under a *weighted* approach, the value of $EFGX_i$ in DIs in which the flowgate is either uncongested or is not invoked is *irrelevant*, since the FGP (the weighting factor) is zero in these DIs. It is in these periods that the RHS and LHS of the constraint equations are not equal and so the LHS measure (which is used) could be a poor estimate of *true* EFGX (based on the RHS).

On this basis, $EFGX^w$ is preferred to EFGX as the measure of flowgate capacity to be used in the incentive scheme, at least conceptually. The practical implications would also need to be considered.

B.4 Flooring the Local Price

B.4.1 Overview

In section 4.3.8, a mechanism was discussed that would prevent the local price paid to a generator (at the margin) being below a specified local price floor (LPF). This was presented in a simple situation where the problematic generator participated in a single congested flowgate. In this case:

$$LMP = RRP - \alpha \times FGP < LPF$$

It is straightforward to revise the local price to equal the LPF by scaling back the flowgate price accordingly:

$$FGP^{revised} = (RRP\text{-}LPF) / \alpha$$

In the general case where the generator participates in multiple flowgates the solution is not so straightforward. In this case:

$$LMP = RRP - \textstyle\sum_k \alpha_k \times FGP_k < LPF$$

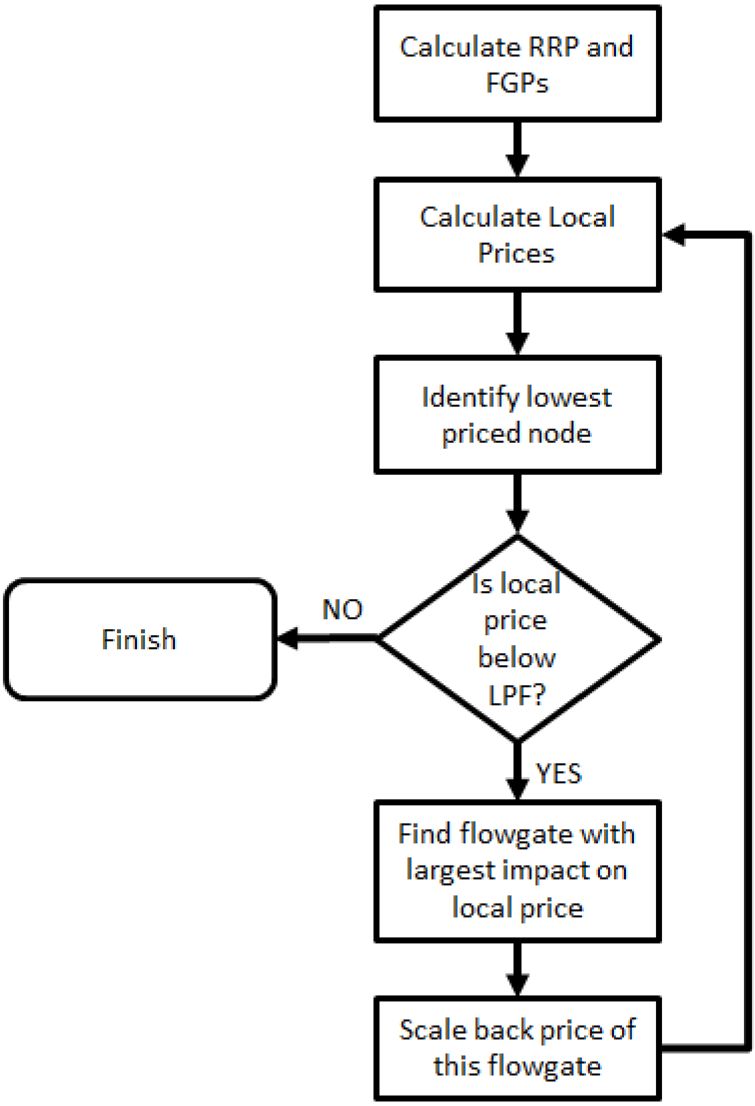
There are many ways that the revised FGPs could be set in order to set the local price at the LPF. For illustration, a couple of options are described below:

1. iteratively revise FGPs that have the largest impact; and
2. scale back all FGPs equally.

B.4.2 Revised largest FGP

The algorithm for revising FGPs is presented in Figure B.6. The first step is to calculate LMPs normally. The node with the lowest LMP is identified. If this is above the LPF, then there is no need for any FGP revision.

Figure B.6 Generalised Flooring Algorithm



Otherwise, the flowgate with the largest impact on the LMP is identified: ie, the flowgate with the highest value of α x FGP. The price on this flowgate is scaled back so that the local price at the lowest-priced node is set to the LPF:

$$LPF = RRP - \alpha_M \times FGP_M^{revised} - \sum_{i \neq M} \alpha_i \times FGP_i$$

Where:

M is the index of the flowgate with the largest impact on LMP.

Therefore:

$$FGP_M^{revised} = \{RRP - LFP - \sum_{i \neq M} (\alpha \times FGP_i)\} / \alpha_M$$

All local prices are then recalculated using this revised FGP and the lowest revised local price is identified. If this is at or above the LPF, the algorithm is finished. If not, the flowgate with the highest impact on this local price is identified and is revised using the same approach as before. The algorithm iterates until it terminates.

The algorithm is likely to converge rapidly. The scaling back of a FGP leads to a rise in local prices for all nodes participating in that flowgate. Therefore, it is likely that only nodes (if any) that do not participate in that flowgate will remain below the LPF after the first iteration.

B.4.3 Scale back all FGPs

In this approach, all FGPs are scaled back together. This is a one-shot approach, without the iteration required in the first approach.

Firstly, LMPs are calculated as usual. A set of flowgates is then defined as follows: a flowgate is in the set if at least one node that participates in the flowgate has a LMP below the price floor.

The prices on *all* of these flowgates are then scaled back by a *common* amount. The scaling factor is set so that it is *just* small enough to bring all of the problematic local prices to LPF or above. This scaling factor can be calculated mathematically, although the maths is not presented here.

This algorithm is fast. However, it may cause *more* flowgate prices to be scaled back – than under the first option – and so impact a broader group of participants. In a sense, it "spreads the pain" compared to the first option.

B.5 Admin Settlements

B.5.1 Administered RRP

Normally, RRP is set equal to the LMP at the RRN. But, under various conditions, RRP is set to a different price: eg, when a price cap or floor applies. This can be expressed mathematically as follows:

$$\text{ROP} = \text{LMP}_{\text{RRN}}$$

Where:

ROP = regional *original* price

During normal periods:

$$\text{RRP} = \text{ROP}$$

During administered RRP periods:

$$\text{RRP} = \text{administered value} \neq \text{ROP}$$

It was demonstrated in A.1.3 that the access settlement payment can be expressed in the form:

$$\text{Access\$} = (A^e - G) \times (\text{RRP-LMP})$$

Where:

A^e is the effective access

The derivation of this concept implicitly assumed that RRP=ROP. In the context of administered RRP, it is more precise to express the formula as:

$$\text{Access\$} = (A^e - G) \times (\text{ROP} - \text{LMP})$$

Therefore, in combination with regional settlement, a generator is paid:

$$\text{Pay\$} = \text{region\$} + \text{access\$}$$

$$= G \times \text{RRP} + (A^e - G) \times (\text{ROP} - \text{LMP})$$

$$= G \times (\text{LMP} + (\text{RRP-ROP})) + A^e \times (\text{ROP} - \text{LMP})$$

Two concerns arise when RRP≠ROP:

1. a generator is no longer paid LMP at the margin and so no-regret dispatch is no longer assured; and

2. a generator is no longer compensated appropriately for being "constrained off" (ie, where A^e>G). The opportunity cost is RRP-LMP but the compensation is paid at ROP-LMP.

B.5.2 Revised Access Settlement

It is proposed that access settlement payments are calculated differently when RRP is administered, based on the formula:

$$\text{Access\$} = (K \times A^e - G) \times (\text{RRP} - \text{LMP})^+$$

The "+" superscript means that any negative value (when RRP<LMP) is set to zero.

In this formula, three revisions are made to the usual calculation:

1. effective access is scaled by a factor, K. K is a regional scaling factor: ie, there will be one value of K for each region. K can be higher or lower than unity;

2. payment is based on the price difference RRP-LMP rather than ROP-LMP. Therefore, the compensation payable is aligned with regional settlement and a generator is paid (in combination with regional settlement) LMP at the margin, where LMP<RRP; and

3. payment is only made where RRP>LMP.

B.5.3 Generator Compensation

The formula ensures that constrained-on generation – with LMP>RRP – is not compensated. This is consistent with access settlement under normal periods. When the administered RRP is due to price capping, and so RRP<ROP, dispatched generators can find themselves constrained-on even in the absence of congestion. For example, a generator with an offer price of $500 who is dispatched against an ROP of $1,000 would find itself constrained-on if RRP is administered at $400.

Generators in this situation are able to claim compensation currently and this arrangement would be expected to continue – and be extended as needed – under OFA. Therefore, similar to the situation under normal conditions, it is not necessary for access settlement to compensate generators for being constrained on, only for being constrained off.

The opposite situation can occur, where a generator that is constrained *off* against ROP becomes *out-of-merit* against an administered RRP. For example, the generator with an offer price of $500 might be constrained off when ROP is $1000. But with RRP capped at $400, the generator is out-of-merit and does not need to be compensated. By comparing LMP to RRP (rather than ROP), the proposed access settlement formula avoids compensating generators in this situation.

If a generator is dispatched and has an offer price and LMP that is *less* than the administered RRP then it will continue to be paid LMP at the margin:

$$\text{Pay\$} = \text{region\$} + \text{access\$}$$

$$= G \times RRP + (K \times A^e - G) \times (RRP - LMP)^+$$

$$= \text{fixed \$} + G \times (RRP - (RRP - LMP))$$

$$= \text{fixed \$} + G \times LMP$$

Therefore, no-regret dispatch is maintained for such generators during administered RRP.

Access Scaling

Access settlement is designed to balance when ROP is used:

$$0 = \sum_i \text{Access\$}_i = \sum_i (A_i^e - G_i) \times (ROP - LMP_i)$$

It is unlikely that access settlement will balance with the revised formula, during administered RRP, if effective access is unchanged:

$$\sum_i (A^e - G_i) \times (RRP - LMP_i)^+ \neq 0$$

Therefore, effective access must be changed to ensure settlement balances. The simplest approach is to scale effective access by a common scaling factor. This is mathematically straightforward and ensures that the "pain" of administered RRP is shared rather than focused on a subset of generators.

Regional scaling factors are preferred to a single NEM-wide scaling factor because RRP administration is done regionally. For example, it is possible that RRP is only administered in a single region and it would be inappropriate then to scale the effective access of generators and DICs in the other regions.

With the scaling of access, the revised formula ensures that:

- firm generators are appropriately compensated for being constrained-off against RRP;
- dispatched generators with LMP<RRP continue to be paid LMP at the margin;
- dispatched generators with LMP>RRP are *not* compensated through access settlement for being constrained on; this will be done through existing mechanisms;
- access settlement balances; and
- access settlement payments are only revised in regions in which RRP is administered.

B.5.4 Setting the scaling factor

For settlement to balance across a region, the aggregate of all the access settlement payments must equal zero.

$$0 = \sum_i \text{access\$}_i = K \times \{\sum_i A^e_i \times (RRP - LMP_i)^+ \} - \sum_i G_i \times (RRP - LMP_i)^+$$

The summation is across all *constrained* generators and importing DICs in the region for which K is being calculated. Unconstrained generators – those that do not participate in any congested flowgates – continue to be excluded from access settlement, even though they could *appear* constrained off if RRP>ROP.[254]

Therefore:

$$K = \sum_i \{G_i \times (RRP - LMP_i)^+ \} / \sum_i \{A^e_i \times (RRP - LMP_i)^+ \} \quad (A2.3)$$

Now A^e_i and (RRP-LMP)+ are generally positive,[255] and so the denominator is always positive, except in the trivial situation where there is no congestion and so access settlement payments are all zero and balance automatically.

254 It is not the role of access settlement to compensate generators that appear constrained off in the absence of congestion, due to RRP administration.

255 Conceivably, A^e could be negative for a generator or DIC subject to mixed constraints.

B.5.5 Flowgate Support Generators

Flowgate support generators have zero access settlement payment under normal conditions. Put another way, their effective access is equal to their dispatched output. It is appropriate that this remains the case under administered RRP. Therefore, the effective access of these generators is not scaled by the K factor. As a result, they need to be excluded from the summation in equation A2.3 above.

The situation is more complicated for a generator or DIC that is subject to mixed constraints: ie, it has positive participation in one congested flowgate and negative participation in another. It is not clear, conceptually or practically, what is the appropriate approach to access settlement for these participants. This needs some further consideration.

B.5.6 Interconnectors

Inter-regional Settlement (IRSR) accrues between regions when the RRPs between the two regions separate, based on the formula (ignoring losses):

$$IRSR = IC \times (RRP_N - RRP_S)$$

Where:

IC is the interconnector flow in a northerly direction

RRP_N is the RRP in the northerly region

RRP_S is the RRP in the southerly region

Obviously, administered pricing in either or both regions will affect the IRSR.

Under access settlement, the IRSR is allocated between the northerly and southerly DICs based on the direction of inter-regional congestion, as described in B.5.6. Similar to the situation with access settlement, the access settlement balances (the two DIC allocations sum to the total IRSR) where RRP is set to ROP in both regions. Where prices are administered in one or both regions, this will not be the case.

Where prices are administered in *one* region only, only the IRSR allocated to the interconnector directed *into* that region should be adjusted. For example, suppose that, in normal regional settlements, the IRSR is allocated to the northerly and southerly interconnectors:

$$IRSR = IC \times (ROP_N - ROP_S) = IRSR_N + IRSR_S$$

Where:

IRSR is the IRSR that would have been calculated in a normal pricing period

$IRSR_N$ and $IRSR_S$ are the IRSR components that *would* have been allocated to the northerly and southerly interconnectors, respectively, in a *normal* pricing period

ROP_N and ROP_S are the original prices in the north and south regions

If the price is administered in the *northerly* region only then define:

$IRSR^A = IC \times (RRP_N - RRP_S)$

$IRSR^A_S = IRSR_S$

$IRSR^A_N = IRSR^A - IRSR_S$

Where:

RRP_N and RRP_S are the reference prices in the north and south regions

$IRSR^A$ is the actual IRSR in the administered price period

$IRSR^A_N$ and $IRSR^A_S$ are the IRSR components allocated to the northerly and southerly interconnectors, respectively, in the final regional settlements

Similarly, if the price is administered in the southerly region only then the reverse process will be used:

$IRSR^A_N=IRSR_N$

$IRSR^A_S= IRSR^A - IRSR_N$

This ensures that DICs that are exporting from the administered region do not have their access settlement payments changed.

If the price is administered in *both* regions then both IRSRs are scaled back by a common factor:

$IRSR^A_N=IRSR_N \times IRSR^A / IRSR$

$IRSR^A_S=IRSR_S \times IRSR^A / IRSR$

B.5.7 FIR Payments

In normal periods, the access settlement payment relating to the firm entitlement of a DIC is allocated between FIR holders, in proportion to the quantity of FIR held. Payment relating to any non-firm entitlement is forwarded to the TNSP in the importing region. A similar approach should apply during administered RRP periods.

C Firm Access Standard

C.1 FAPS Conditions

C.1.1 Overview

This section presents an *example* of FAPS conditions. This example is intended to illustrate both the scope of what would need to be specified and the principles that could inform the specification. The example is presented in Table C.1, below. It should be emphasised that this is for illustration only. FAPS conditions would be specified during OFA implementation or in the early period of OFA operation.

Recall from section 5.2.2 that FAS conditions would be designed to reflect the annual snapshot in which firm access is most likely to be most valuable. In the example this is considered to be:

- under summer weather conditions;
- during regional peak demand; and
- when regional prices are high.

Table C.1 Example specification of FAPS conditions

Variable	Specification
Scheduled Generation	Expected peak-period availability
Semi-scheduled	Expected peak-period UIGF
NSG Non-intermittent	Expected peak-period availability
NSG Intermittent	*Legacy Intermittent*: Expected peak-period availability *Other Intermittent*: Expected peak-period UIGF
Transmission	System Normal: all assets in service Ratings based on peak demand ambient conditions
Demand	Scheduled Demand: not consuming Non-scheduled Demand: regional peak demand
Interconnectors	Regulated ICs do not provide flowgate support, so status irrelevant Basslink: max flow into Victoria

Note: NSG means "non-scheduled generation".

UIGF means “unconstrained intermittent generation forecast”.

These conditions are discussed in turn below.

C.1.2 Non-intermittent Generation

Assumptions for scheduled generation are relevant only in relation to estimating flowgate support. The assumed level of output is independent of registered access: ie, firm and non-firm generators are equally likely to provide flowgate support.

Since both prices and demand are high, it is assumed that all generation operates at its estimated availability. Availability, in turn, would be based on both generation capacity and on the peak availability factor.

The same assumption is used for conventional (non-intermittent) non-scheduled generation which would similarly be assumed to operate at full availability during high price periods.

C.1.3 Intermittent Generation

Different assumptions are made for:

- semi-scheduled generation;
- *legacy intermittent NSG*: non-scheduled intermittent generation that was registered before the introduction of the semi-scheduled generation class; and
- *other intermittent NSG*: non-scheduled intermittent generation registered after this date.

Different rules apply to the registration of intermittent generation before and after the introduction of semi-scheduled generation. In the later period, larger intermittent generators must be registered as semi-scheduled; in the earlier period, they would be registered as non-scheduled.

Semi-scheduled generators can, when operating, only provide flowgate support. On the other hand, non-scheduled generators, when operating, can detract from flowgate capacity. The impact of the legacy intermittents can, therefore, substantially impact on TNSP FAPS and FAOS obligations.

Consider a 100MW, legacy intermittent windfarm. Its expected technical availability at peak – ignoring intermittency – might be 95MW, say. Its expected UIGF – which takes into account both technical availability and wind resource – might be only 40MW, say. Its output, operationally, could be as high as 100MW, of course. If FAPS conditions used the 40MW figure, windfarm output could operationally be up to 60MW higher than the FAPS output, meaning that FGX operationally could be 60MW lower than FAPS capacity: if the windfarm was located behind the relevant constraint and the wind was blowing strongly. This would cause firm entitlements to be scaled back commensurately. If, on the other hand, FAPS conditions used the 95MW figure, there is a much lower risk from shortfalls being caused by the intermittent generation. For this reason, the example FAPS condition is based on the 90MW figure: the expected peak *technical* availability.

A *non*-legacy non-scheduled windfarm is unlikely to be more than 30MW in size (any larger than this it would be required to be semi-scheduled), so even if the higher, technical availability were used under FAPS conditions, the associated risk of shortfalls is relatively low.

Allocated Transitional Access (TA) will be scaled back in order to be accommodated within the FAPS capacity level at the time of OFA implementation. Therefore, the lower FAPS capacity implied by this treatment of the legacy intermittent will not prompt any additional network expansion. Rather, it will mean that affected generators will be allocated a lower amount of TA.

C.1.4 Transmission

Transmission is assumed to be system normal, with all assets in service. This means that any outages – planned or unplanned – could lead to shortfalls. The incentive scheme will provide incentives on TNSPs to minimise the frequency, duration and impact of such shortfalls.

Transmission ratings are based on peak demand conditions. In all mainland regions, this means *summer* conditions.[256] This ensures consistency between the demand and transmission conditions.

C.1.5 Demand

Since RRPs are assumed to be high, it is assumed that all scheduled demand will be "off": ie, not consuming power.

Non-scheduled demand will be based on regional peak demand. This would need to be defined precisely: eg, what point-of-exceedance would be used and what weather conditions would be assumed for weather-dependent demand.

C.1.6 Interconnectors

Regulated interconnectors (ie DICs) are scheduled and do not provide flowgate support.[257] Therefore, they have no effect on EFGX[258] and so there is no necessity to include them in the specification of FAPS conditions.

256 OFA is not expected to be introduced in Tasmania, at least in the first instance, and so FAPS conditions are not relevant for that region.

257 Except under operational conditions in which EFGX is negative. FAPS capacity must equal or exceed TFGX, which is non-negative, by definition. Therefore it can be assumed that negative EFGX will never occur under FAPS conditions and so DICs never provide flowgate support.

MNSPs are treated as scheduled generators in the importing region. Their treatment under FAPS conditions should therefore be consistent with that for scheduled generation. If the Victorian RRP is high, it is reasonable to assume that Basslink would be flowing into Victoria, at the level of its expected peak availability.

C.2 Nested Caps in the Incentive Scheme

C.2.1 Identified versus Inferred Operating Conditions

The operational incentive scheme aims to apply differently structured penalties depending upon the type of condition causing flowgate shortfall costs. This is done *implicitly* rather than explicitly. The same rules are applied under all conditions: system normal, planned outages, forced outages and so on. However, because these conditions have different intrinsic characteristics, they interact with the scheme's nested caps in different ways and place differing incentives on the TNSP as a result.

This is best explained through illustrative examples. Three example conditions are considered in turn below: a planned outage, a forced outage and system normal. TNSP penalty outcomes are derived, based on an example scheme with nested caps specified below:

- annual cap: $10m;
- daily cap: $200,000 (one 50th of the annual cap); and
- trading interval cap: $20,000 (one 10th of the daily cap) on any flowgate.

C.2.2 Planned Outage

The characteristics of, and objectives for, planned outages are understood to be as follows:

1. they typically have an extended duration, from several days to several weeks;
2. advance notice of planned outages to market participants is generally possible and desirable;
3. they should be scheduled for periods when congestion costs are likely to be low; and
4. they should be cancelled, where practical, if conditions change adversely from those expected.

Consider a planned outage with the following characteristics:

---

258 Although plausibly they could have an impact, if there are some non-linear interconnector terms included in the RHS of a constraint equation.

- it is of six weeks duration;
- it reduces FGX by 1000MW below TFGX on a particular flowgate; and
- it gives rise to an expected FGP of $5/MWh on that flowgate for 10 hours per business day.

Maximum penalties under the incentive scheme would then be:

- $2,500 (1000MW x $5/MWh x 0.5) per TI: this does not hit the TI cap;
- $50,000/day (20 x TI penalty), which does not hit the daily cap; and
- $250,000/week (5 x daily penalty) and so $1.5m for the 6-week duration.

No incentive scheme caps are hit and so shortfall costs are charged to the TNSP *in full*. The TNSP is fully responsible for, and exposed to, congestion costs. Leaving aside system normal and forced outage penalties, a TNSP could have *six* such outages in a year before hitting the annual cap.

Therefore, a TNSP has a strong incentive to efficiently reduce its exposure to penalties by:

- rescheduling the outage to a period with a lower expected FGP;
- shortening the outage duration: eg, by overnight or weekend working; and
- giving generators advance notice: possibly encouraging them to align their own outage plans or otherwise to change operating or trading plans to reduce congestion costs. (Note that firm generators are not exposed to congestion costs in this example, but non-firm generators might be).

In summary, the typical characteristic of the planned outage – and the design of the caps – means that a TNSP is likely to have a *high* exposure to the consequential shortfall costs.

C.2.3 Forced Outage

Next, a forced outage is considered. It is assumed to occur in a peak period on a major flow path and so create severe congestion. Its assumed characteristics are as follows:

- it reduces FGX by 1000MW below TFGX on a particular flowgate; and
- it creates a FGP of $1000/MWh on that flowgate, which remains high until the failed element is restored.

Penalties under the incentive scheme are then:

- the shortfall cost is $500,000 per TI: the TNSP penalty is therefore capped at $20,000 per TI;

- this continues until the daily cap of $200,000 is hit (after 5 hours) or the element is restored;
- this repeats the following day, and so on, until the element is restored; and
- the annual cap will only be hit if the forced outage continues for 50 days.

A TNSP has a strong incentive to ensure the element is returned within five hours. If it does not achieve this, the incentive is then to return the element before the next day. And this incentive keeps repeating, day after day, for a maximum of 50 days.

The TNSP also has an incentive to reduce the frequency of forced outages. It is recognised that in the above example, the TNSP is exposed to only a small percentage (4%) of the estimated shortfall cost. However, there will be other forced outages during less stressful conditions when the percentage exposure will be higher.

Because any severe congestion caused by a forced outage will cause the TI cap to be hit, the incentive scheme penalty is similar to a tariff: $20k for each TI in which a major forced outage occurs and then $200,000 for each day it continues. This is not dissimilar – in structure – to the existing STPIS incentive on forced outages.

C.2.4 System Normal

A third possible example is of flowgate shortfalls occurring during system normal periods. This may be due to a planning failure: for example due to a TNSP deliberately delaying a planned expansion in order to reduce capital expenditure costs. Alternatively, it might be because flowgate capacity off-peak is below the (peak) FAPS level.

The assumed characteristics are:

- a relatively low FGX shortfall of 100MW on a flowgate; and
- a modest average FGP of $2/MWh on that flowgate.

On these assumptions, penalties are then:

- $100 shortfall cost per TI, which does not hit the TI cap;
- $4,800 shortfall cost per day, which does not hit the daily cap; and
- $1.7m per year, which does not hit the annual cap.

In this case, the TNSP could be fully exposed to the shortfall costs (depending upon what other incentive penalties accumulate during the year) and will have an efficient incentive to undertake the necessary capex or otherwise ameliorate the situation. Of course, under more severe assumptions, the annual cap will be hit and the degree of incentive reduced.

C.2.5 Discussion

The above examples illustrate how the incentives on a TNSP under the incentive scheme may vary depending upon the underlying conditions causing flowgate shortfall costs. For a planned outage, a TNSP is very sensitive to expected FGP and will either seek outage periods where FGP is likely to be low or aim to minimise duration and FGX impact. For a forced outage, a TNSP instead aims to reduce average forced outage frequency and duration, although it can perhaps respond to severe outages in order to reduce the duration of that particular outage. For a system normal shortfall, a TNSP may be incentivised to undertake capex.

The assumptions presented above are illustrative only. In practice, typical outage characteristics may vary substantially from those presented. The parameters of the real scheme will be tuned to actual outage characteristics, based on some quantitative and historical analysis.

It should also be noted that the examples above involve only a single congested flowgate. In practice, multiple flowgates will bind over a period. The TI cap applies to each individual flowgate, but the other caps apply in aggregate across all flowgates.

C.3 Shortfall Cost Allocation

C.3.1 Overview

To settle the incentive scheme, there is a need to attribute capped shortfall costs between firm participants. This can be done by introducing scaling factors such that scaled shortfall costs and capped shortfall costs are identical. The scaling factors can then be applied to the actual shortfall costs incurred by each participant.

C.3.2 Scaling Factors

Suppose that the incentive scheme has TI, daily and annual caps. Define the following variables:

CAP = scheme nested cap

s = scaling factor

c = (uncapped) shortfall costs for each TI and flowgate

C = aggregate shortfall costs

Let the subscripts i, k and d be indices for TI, flowgate and day, respectively. Let superscripts T, D and Y refer to TI, daily and annual parameters, respectively.

Define the TI scaling factor as:

$$s^T_{ik} = \min(1, CAP^T/c_{ik})$$

Then:

s^T_{ik} x c_{ik} =min(c_{ik} ,CAP^T) = TI-capped shortfall cost

So, the scaled TI shortfall cost is the same as the capped shortfall cost.

Now, the daily cap only applies if the total of the TI-capped shortfall costs over a day is greater than the daily cap.

Total of TI-capped shortfall costs over day = $C^D_d = \sum_i \sum_k s^T_{ik} \times c_{ik}$

Where the summation is over all TIs in the day, d, and across all flowgates.

Define the daily scaling factor as:

$s^D_d = \min(1, CAP^D / C^D_d)$

Then:

$s^D_d \times C^D_d = \min(C^D_d, CAP^D)$ = daily-capped shortfall cost

Finally, calculate the annual shortfall cost and scaling factor similarly.

Total of daily-capped shortfall costs over year $C^Y = \sum_d s^D_d \times C^D_d$

Where the summation is over all days in the year.

Define the annual scaling factor as:

$s^Y = \min(1, CAP^Y / C^Y)$

Then:

$s^Y \times C^Y = \min(C^Y, CAP^Y)$ = annual capped shortfall cost

Now define the *scaled shortfall cost* c^s_{ik} as:

$c^s_{ik} = s^T_{ik} \times s^D_{d(i)} \times s^Y \times c_{ik}$

Where d(i) is the particular day in which the trading interval, i, falls. So, the scaled shortfall cost is the original shortfall cost scaled by all three scaling factors.

The scaled shortfall cost complies with all the nested caps in the scheme. This is because:

- $s^T_{ik} < 1$ only if the TI cap is hit, for trading interval i and flowgate k, in which case the shortfall cost is reduced to the TI cap;
- $s^D_d < 1$ only if the daily cap is hit on day, d, in which case the total shortfall cost on that day is reduced to the daily cap; and

- $s^Y < 1$ only if the annual cap is hit, in which case the total shortfall cost on that day is reduced to the daily cap.

In the example there are three nested caps. However, the algebra can easily be extended to any number of nested caps: eg, including also weekly and monthly caps. The scaling factors are determined for each timescale in turn, starting with the shortest timescale and working outwards progressively.

C.3.3 Participant shortfall costs

Shortfall costs are shared between participants in proportion to their target firm entitlements as follows.

Let participant j have registered access level Aj and participation αjk in flowgate k (for simplicity, it is assumed that the participation factors do not vary by time). Then, the target entitlement is:

Target entitlement = $TE_{jk} = A_j \times \alpha_{jk}$

The actual entitlement in trading interval i is:

Actual entitlement = $E_{ijk} = K_{ik} \times A_j \times \alpha_{jk}$

Where K is the firm entitlement scaling factor.

$K_{ik} = EFGX_{ik}/TFGX_k$

It is assumed that $EFGX_{ik} < TFGX_k$. If $EFGX_{ik} \geq TFGX_k$ then there are no shortfall costs, individually or in aggregate, and the algebra is trivial.

The shortfall cost incurred by a participant is the *value* of the difference between the target and actual entitlements. Value is entitlement multiplied by FGP.

Participant shortfall cost = $cp_{ijk} = (TE_{jk} – E_{ijk}) \times FGP_{ik} = (1- K_{ik}) \times A_j \times \alpha_{jk} \times FGP_{ik}$

The total shortfall cost incurred by all participants is then:

Total shortfall cost = $\sum_j cp_{ijk}$

$= \sum_j (1- K_{ik}) \times A_j \times \alpha_{jk} \times FGP_{ik}$

$= (1- K_{ik}) \times FGP_{ik} \times \sum_j \times A_j \times \alpha_{jk}$

$= (1- K_{ik}) \times FGP_{ik} \times TFGX_k$

$= (TFGX_k – EFGX_{ik}) \times FGP_{ik}$

This is the same as the shortfall cost used in the incentive scheme, ie:

$c_{ik} = \sum_j cp_{ijk}$

Therefore the total of the actual shortfall costs borne by participants equals the shortfall cost referenced in the incentive scheme. Of course, this is the purpose of defining shortfall cost in this way.

The capped shortfall cost attributed to each participant is then calculated by multiplying the participant shortfall cost by the scaling factors calculated earlier:

Capped participant shortfall cost = c^{sp}_{ijk}

$= s^T_{ik} \times s^D_{d(i)} \times s^Y \times c^p_{ijk}$

The total capped shortfall cost attributed to each participant for the year is then calculated by summing these individual values:

$C^Y_{pj} = \sum_{ik} c^{sp}_{ijk}$

$= \sum_{ik} s^T_{ik} \times s^D_{d(i)} \times s^Y \times c^p_{ijk}$

By definition, these attributed shortfall costs will sum to the overall shortfall cost:

$\sum_j C^Y_{p}j = \sum_j \sum_{ik} s^T_{ik} \times s^D d_{(i)} \times s^Y \times c^p_{ijk}$

$= \sum_{ik} s^T_{ik} \times s^D_{d(i)} \times s^Y \times \sum_j c^p_{ijk}$

$= \sum_{ik} s^T_{ik} \times s^D_{d(i)} \times s^Y \times c_{ik}$

$= C^Y$

Therefore, the attributed participant capped shortfall costs:

- in aggregate, comply with the nested caps in each timescale; and
- across the year, sum to the total capped shortfall cost.

Therefore, these can be used as the basis for settlement of the incentive scheme.

D Access Pricing

D.1 Meshedness

D.1.1 Overview

The concept of meshedness is introduced in section 6.2.4. It is informally referred to there as the number of transmission paths in parallel with an element. This appendix defines meshedness more formally.

In the network topology, each network element has two defined end nodes. If a power source (ie, a generator) is placed at one end node and a power sink (ie, a customer load) of equal size is placed at the other end node, power will flow through the network as a result. Some power will flow through the element itself; the remainder will flow along paths that are parallel to the element: ie, have the same start and end node.

The meshedness of an element is defined as:

> Meshedness = total power flowing through network / power flowing through the element

For example, if the source and sink had a power of 100MW and just 25MW flowed through the element, the meshedness of the element is 4. The lowest possible value of meshedness is one: when the element is a radial element with no parallel paths.

D.1.2 Distribution Factors

Define a distribution factor d_{ik} which is the MW flow through line i when the 1MW power source and sink is placed across the ends of line k. Refer to the start and end nodes of line k as s(k) and e(k), respectively.

Distribution factors are mathematically related to participation factors. Recall that a participation factor α_{ik} is the power that flows through element k when a 1MW generator at node i supplies a 1MW load located at the RRN.

Consider three different load flows:

- load flow one: a 1MW generator at node s(k) supplies a 1MW load at the RRN;
- load flow two: a 1MW generator at the RRN supplies a 1MW load at e(k); and
- load flow three: a 1MW generator at node s(k) supplies a 1MW load at node e(k).

Load flow three is just load flow one and load flow two superimposed. Therefore, the flow on each element in load flow three is the sum of the flows from load flows one and two.

Load flow three is the one which determines distribution factors:

d_{ik} = flow on element i in load flow 3

Flows on element i in the other load flows can be calculated from the participation factors:

Flow on element i in load flow 1 = $\alpha_{is(k)}$

Flow on element i in load flow 2 = -1 x $\alpha_{ie(k)}$

Therefore:

$$d_{ik} = \alpha_{is(k)} - \alpha_{ie(k)} \qquad \text{(A4.1)}$$

D.1.3 Meshedness factors

Define:

μ_k = meshedness of element k

Then, by the definition of meshedness:

$$\mu_k = 1/\ d_{kk}$$

From equation A4.1 above:

$$\mu_k = 1/\ (\alpha_{ks(k)} - \alpha_{ke(k)})$$

Therefore, meshedness factors can easily be calculated from the participation factors.

D.2 Replacement

D.2.1 Overview

Section 6.2.2 noted that the expected end-of-life of each network element would be defined in the network topology used in the access pricing model. The capacity of the element would be removed in the end-of-life year, leading to possible thermal overload in the firm generation dispatch load flow, which prompts an expansion in the baseline and/or adjusted expansion plan.

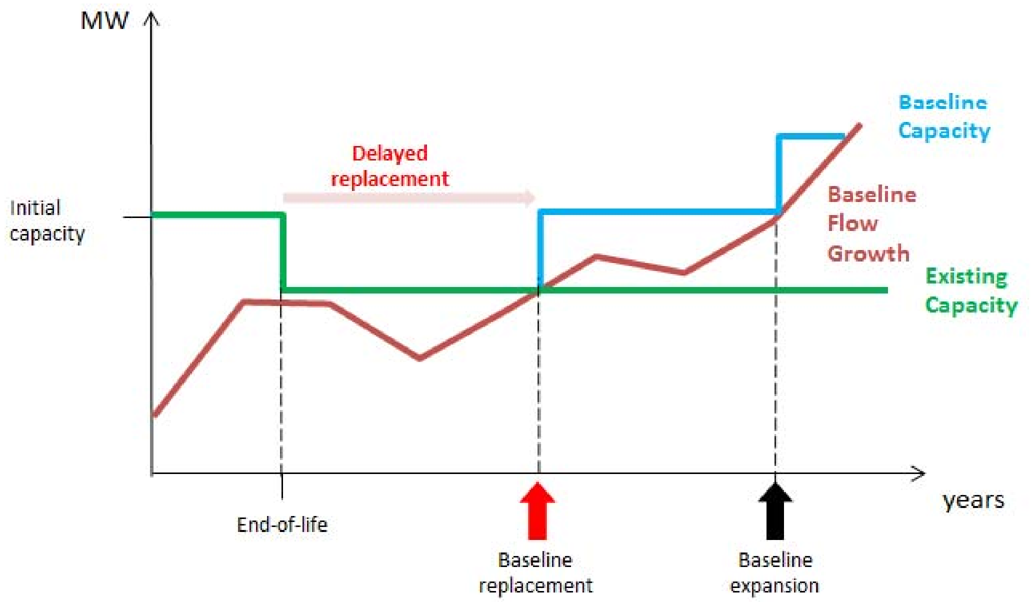
Such an expansion can be considered to be a *replacement* of expiring existing capacity rather than the addition of new capacity. The pricing model does not distinguish between these alternatives: the expansion cost and size is assumed to be the same in each case. However, other NEM processes – eg, the RIT-T– *do* make the distinction.

Of course, if there were no end-of-life in the pricing model – meaning that network assets are implicitly assumed to last for ever – there would correspondingly be no modelling of replacement.

D.2.2 Delayed Baseline Replacement

Figure D.1 (Delayed Baseline Replacement) presents a possible scenario in which replacement of a network asset occurs, but not immediately: replacement is *delayed*. This occurs because there is enough spare capacity on the element such that at end-of-life, even with reduced network capacity[259], there is still sufficient capacity and no immediate replacement is needed.

Figure D.1 Delayed Baseline Replacement



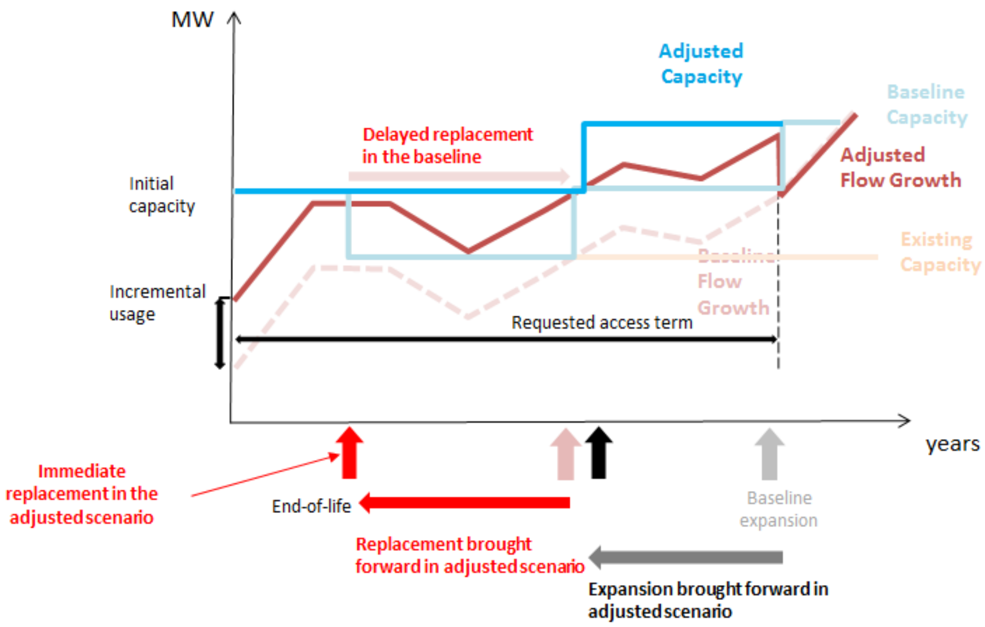
Subsequent growth in the baseline flow on the element means that replacement is eventually required. A later baseline expansion is also needed as the flow grows further.

D.2.3 Advanced Adjusted Replacement

Figure D.2 (Advanced Adjusted Replacement) presents the corresponding adjusted scenario. The incremental usage from the access request means that replacement can no longer be delayed: it is required immediately. Therefore, compared to the baseline scenario, the replacement is advanced. The cost of this advancement will be included in the access price.

259 by implication, the capacity on the element is provided by multiple assets and only one asset has reached end-of-life. So, the element capacity is reduced but does not fall to zero

Figure D.2 Advanced Adjusted Replacement



The later baseline expansion is also advanced in the adjusted scenario.

D.2.4 Conclusions

The modelling of network assets as having finite life means that replacements are modelled in the pricing model, although these are not explicitly distinguished from expansions. Replacement costs will be included in the access price where they are different between the baseline and adjusted scenarios. This is likely to occur where a replacement can be delayed (or, possibly, is not required at all) in the baseline scenario and must be made immediately (or at least earlier than in the baseline) in the adjusted scenario.

Delayed replacement in the baseline is most likely where there is substantial spare capacity at the time of asset end-of-life. This is most likely when there is a flat or declining demand for firm access in a particular zone.

E Issuance

E.1 Long-term Intra-regional

E.1.1 Sellback at LRDC

As discussed in section 7.2.2, generators have the right to sell back their long-term intra-regional firm access at any time, at the current Long Run Decremental Cost (LRDC). It is asserted that if, hypothetically, a sellback was made immediately after a purchase, the sellback and purchase prices would be equal and so the transactions would exactly net out. This section demonstrates that assertion, at least in relation to the thermal cost.[260]

The access price is calculated as an LRIC:

Access price = LRIC = adjusted expansion cost – baseline expansion cost

The expansion cost for a scenario is a function of the existing network topology (ENT) and the scenario forecast:

Expansion cost = EC(scenario forecast, ENT)

Therefore:

LRIC = EC(baseline forecast + access request, ENT) – EC(baseline forecast, ENT)

With LRDC, the access request is instead *subtracted* from the baseline forecast.

LRDC = EC(baseline forecast, ENT) – EC(baseline forecast-access request, ENT)

In the situation considered, the sellback is made immediately after the access request has completed. Thus, the baseline forecast will have been updated to reflect the completed request. The ENT, on the other hand, will *not* have changed:

Sellback price = EC(baseline forecast + access request, ENT) – EC(baseline forecast + access request – access request, ENT)

= EC(baseline forecast + access request, ENT) – EC(baseline forecast, ENT)

= access price

Therefore:

Sellback price = access price

260 As discussed in section 6.2.5 stability costs are not priced using the LRIC methodology and so LRDC is not clearly defined.

For this equation to hold, the sellback must occur immediately following the request completion. Specifically:

- there must be no other changes to the baseline: eg, from some other access requests having completed in the interim; and
- there must be no changes to the existing network topology: eg, due to the TNSP planning expansions to accommodate the access request.

E.1.2 Sequencing of Request Processing

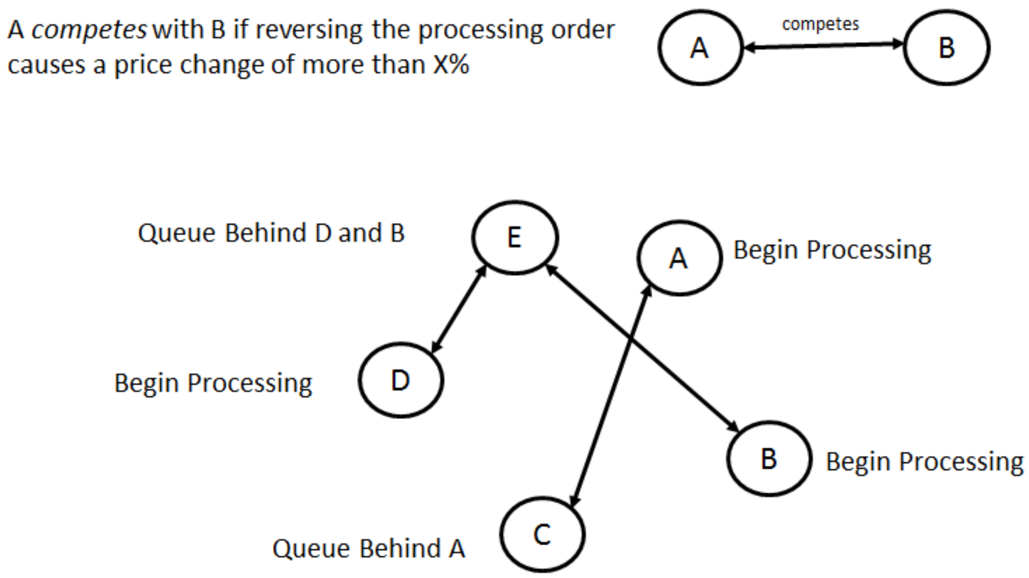
As discussed in section 7.2.2, concurrent access requests will be ordered in accordance with a specified queuing policy. Sequencing will be based on this order.

Requests can be processed in parallel where they do not *compete*: ie, the completion of one request will not change the price of another request. Therefore, sequencing can be undertaken in accordance with the following principles:

- non-competing requests can be processed in parallel; and
- competing requests are processed sequentially, in accordance with the queuing order.

This is illustrated in Figure E.1. Five requests are shown and are labelled in order of queueing: ie, A, B, C, D and E. The arrows connecting pairs of requests indicates where requests compete: eg, A competes with C.

Figure E.1 Queuing Competing Requests



The sequence is then as follows:

1. A commences processing immediately, since it is at the head of the queue;

2. B can also commence processing immediately, since A and B do not compete: if there were TNSP resource constraints on processing the two in parallel, A would take precedence;

3. C is queued behind A: it cannot commence processing until A has finished;

4. D can commence processing immediately: it does not compete with A or B or C; and

5. E competes with B and D and cannot commence processing until B and D have finished.

In general, suppose that X is an access request. Define S^X to be the set of all currently queued access requests that compete with X and that are ahead of X in the queue. Then X cannot commence processing until *all* of the access requests in S^X have been processed.

E.2 Long-term Inter-regional

E.2.1 Overview

As discussed in section 7.2.3, long-term inter-regional access would be issued through an auction process based on a set of principles which are reproduced below:

- bidders can submit separate annual bids for FIRs, but cleared amounts must, in aggregate have a strip shape;
- a reserve price is included which ensures that auction revenue equals or exceeds the access price of the inter-regional access that is issued;
- the sequencing of inter-regional auctions (on different DICs) and intra-regional access request processing is clearly defined;
- FIR holders can offer to sell some of their holdings through the auction, for purchase by other participants; but TNSPs are not permitted to re-purchase FIRs; and
- the auction is cleared on a platform that is developed and operated by AEMO but then settled directly between auction participants and TNSPs.

This section presents, for illustration, two possible high-level designs for the auction which would conform to these principles. It is proposed that a pair of DICs is sold in a combined auction, but this creates some additional issues, as discussed in section 7.2.3. For simplicity, therefore, it is assumed here that access on each DIC is auctioned separately.

The two designs are:

- a static auction; and

- a dynamic auction.

These are discussed in turn.

E.2.2 Static Auction

Overview

Static auctions will be familiar to NEM participants as they are the auction format that is used in the NEM: for example, in dispatch and in the settlement residue auction. Bids and offers are submitted to the auctioneer (AEMO in each case, and also for the inter-regional firm access auction). Using an auction clearing engine, based on some established clearing principles and objectives, a *clearing point* is determined. The clearing point specifies:

1. the quantity purchased and/or sold by each auction participant; and

2. the clearing price payable for each product.

For example, in NEM dispatch, the quantities are the dispatch targets and the clearing prices are the dispatch prices.

The auction is *static*, because it is undertaken in a *single* round. Bidders have no opportunity to revise their bids *dynamically* over several auction rounds.[261]

Averaging of Bids and Access Price

The reserve price requirement specifies that the auction revenue must equal or exceed the access price of the issued access. The auction revenue is a function of the auction volume and the clearing price. Clearing prices will be set for each year of the issued access term. Clearing volume, on the other hand, will be constant, since the issued access must be a strip. Therefore, auction revenue is:

$$\text{Auction revenue} = \sum_y (Q \times P_y) \geq \text{access price} \qquad \text{(A5.1)}$$

Where:

Q is the cleared volume

P_y is the clearing price in year y

Dividing both sides of inequality A5.1 by the quantity and the number of years gives:

$$\sum_y P_y / N \geq \text{access price}/(Q \times N) \qquad \text{(A5.2)}$$

261 Although generators have the opportunity to revise their bids for *future* dispatch intervals in the light of earlier dispatch outcomes, so NEM dispatch has some of the characteristics of a dynamic auction.

The LHS of equation A5.2 is the average clearing price. The RHS will be referred to as the *average access price*.

In the auction, as is conventional, the highest bids, in each year, will be cleared first and the clearing price will be set to the price of the marginal bid: ie, the lowest-priced bid cleared, at least in part. Since this is true in each year, it is also true on average, ie:

average clearing price = average marginal bid

Where, again, the averaging is over the N years of the issued access term.

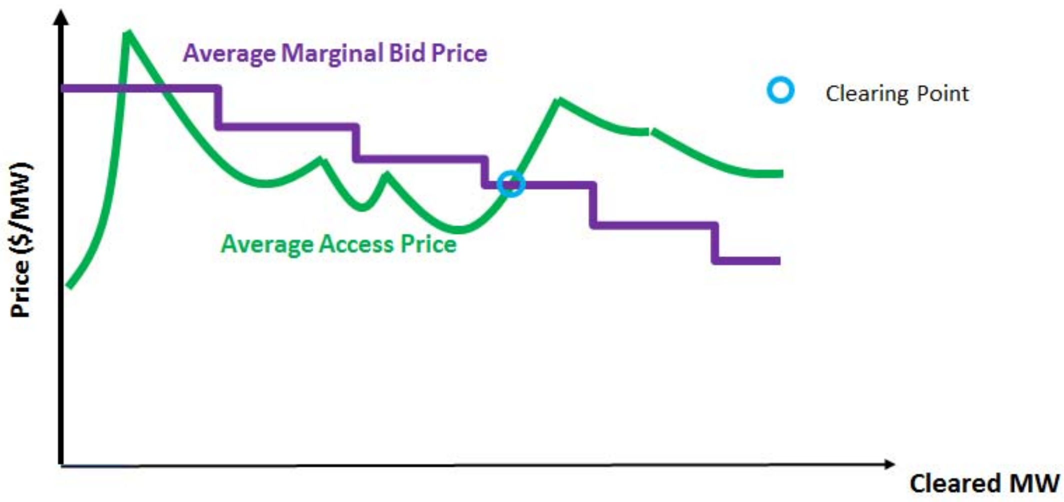
Therefore, the reserve price constraint can be expressed as:

Average marginal bid ≥ average access price

Reserve Price Constraint

The reserve price constraint is presented graphically in Figure E.2. The average marginal bid curve is downward sloping: by definition, because the highest-price bids are cleared first in each year. The average access price is "wavy": it is piece-wise convex and will typically change from upward sloping to downward sloping at various *breakpoints*. These breakpoints represent points at which additional expansion is triggered in one of the auction years.

Figure E.2 Reserve Price Constraint



The graph presents these curves only as a function of the cleared MW, Q, implicitly assuming that the number of years in the access term, N, is fixed. In practice, both N and Q can be varied, and so the two curves are actually three dimensional surfaces rather than lines. Obviously, this is complicated to present in a graph.

Clearing Point

The reserve price constraint defines a range of possible clearing points: in the figure, the clearing point could be anywhere on the marginal bid curve, where it is above or on the average access price curve. Some further principles or objectives need to be included to specify a unique clearing point.

It is proposed to select as the clearing point the point with maximum MWh that satisfies the reserve price constraint. This point is indicated on Figure E.2.

The reasons for this approach are:

1. it is simple and transparent: ie, the clearing point is clearly seen from the graph of bid curve against average access price;

2. it supports inter-regional expansion; and

3. it is, in practice, likely to give a clearing point that maximises value.

This last point is addressed in the next section.

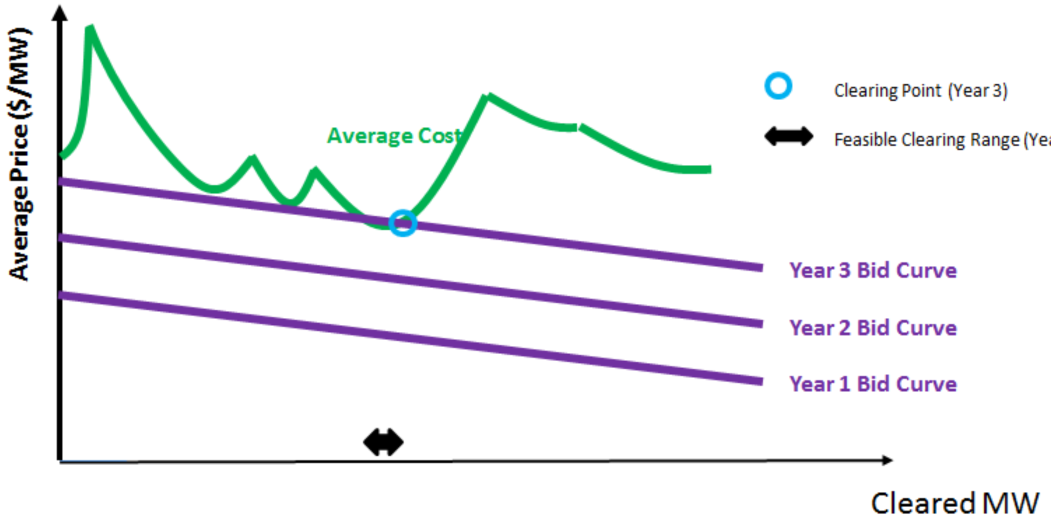
Maximum value

Conventionally, auction design selects a clearing point as the point at which the auction value is maximised. In the context of the access auction, the auction value is the value of the firm access to purchasers minus the cost of firm access provision to the TNSP.

The point of maximum volume is unlikely to maximise value. However, since the reserve price constraint must be satisfied, it is known that the maximum value clearing point must be somewhere in the feasible range. In the figure above, this range is large, but there are reasons to think it might be quite narrow in practice, for reasons discussed below and illustrated in Figure E.3.

In the presented example, it is assumed that, in the first year that the auction takes place, demand for inter-regional firm access is low and there is no clearing point that satisfies the reserve price constraint. Over subsequent years, demand increases and eventually, in year 3 in the figure, there is a range of clearing points over which the reserve price constraint is satisfied. (For simplicity of illustration, it is assumed that the access price curve remains constant.)

Figure E.3 Auction Clearing and Demand Growth



Because of the assumed *gradual* demand growth, the feasible range is quite small. Thus, by definition, the difference between the maximum value clearing point and the maximum volume clearing point is also small.

E.2.3 Dynamic Auction

Overview

In a dynamic auction, bids that are submitted in the first round lead to some *indicative*, rather than final, clearing point. Participants are then permitted to revise their bids, in the light of this new information, leading to a second round clearing point being established and notified to participants. This process continues until some specified termination conditions are met, at which point the process completes and the auction clears at the clearing point established in the final auction round.

An inherent advantage of dynamic auctions over static auction is that the market value of the auctioned product is progressively revealed, reducing concerns that participants may have about the *winner's curse*, in which the bidder who over-estimates the value of the product ends up winning the auction and buying the product at an uncompetitive price. To avoid this, bids may be quite conservative in a static auction. In the OFA context, this might lead to a static auction failing to clear, even where there is sufficient, *latent* demand to support interconnector expansion.

Clock Auction

The design proposed is based on a clock auction. A "clock" generically represents an individual auctioned product. In the firm access context, each year of firm access on the auctioned DIC would be represented by a separate clock.

Unlike in the static auction, where the clearing price is determined by the intersection of auction supply and demand, in a clock auction the price for each clock is set administratively prior to each auction round. Bidders then submit *quantity* bids, for each clock, expressing the quantity they would choose to purchase at that price.

After each round, the total bid quantity for each clock is compared to the supply that is being auctioned. Where demand exceeds supply, the clock price is increased incrementally. A new auction round then takes place. This continues until demand matches supply on each clock. The auction then terminates and clears, based on the latest clock price and bid quantities.

Reserve Price Constraint

After each auction round, a *provisional issuance* is determined. This would be a strip of inter-regional firm access on the auctioned DIC, with quantity Q and term N, which satisfies the following conditions:

- potential auction revenue equals or exceeds the access price; and
- subject to the first condition, issued volume (N × Q) is maximised.

Potential auction revenue means the auction revenue that could be received if the issuance occurred, based on the latest bids. It is defined as:

$$\text{Potential auction revenue} = \sum_{y=1\text{to}N} [P_y \times \min(Q, BQ_y)]$$

Where:

P_y is the clock price for year y

BQ_y is the bid quantity for year y

It is possible that Q=0, meaning that both potential auction revenue and access price equal zero, and so the reserve price constraint is satisfied. Indeed, this is likely in the early auction rounds, as low clock prices will be set initially.

Iteration

The clock price is increased for the year y clock where the bid demand BQ_y exceeds the quantity of the provision issuance in that year. That would be either because $BQ_y > Q$ or because y>N. For clocks in which there is no excess demand there is no price change.

The auction is then re-run. A provisional issuance is recalculated. It would be necessary to ensure that neither N nor Q *decreases* from one year to the next.

The auction terminates when $BQ_i \leq Q$ for every clock. Potentially, this could be when $BQ_i = Q = 0$ in every clock: ie, there is insufficient demand to prompt any issuance.

A key design aspect for clock auctions is to decide by how much to increment the clock prices. If the increment is too small, there will need to be many auction rounds before the auction terminates and this may create practical difficulties and costs. On the other hand, if the increment is too large, a potential clearing point may be missed, due to the increased price causing bid quantity to fall substantially. A second design issue is what restrictions to place on bids. Typically, bid quantities from individual bidders are not permitted to *increase* from one round to the next. This ensures that bidders must participate fully in the auction from the first round. It also helps with auction convergence.

E.3 Short-term Auction

E.3.1 Auction Formulation

Overview

Short-term issuance is limited by the *FAPS constraints* that require that TFGX, post-auction, is no higher than the existing *FAPS capacity*: ie, effective flowgate capacity under FAPS conditions. The FAPS capacity would be calculated for the year in which the particular short-term firm access product was to be issued.

The post-auction TFGX is defined by the formula:

$$TFGX_k = \sum_i \alpha_{ik} \times R_i \quad (A5.3)$$

Where:

R_i is the post-auction registered access of firm participant i

α_{ik} is the positive participation[262] of firm participant i in flowgate k

Therefore the FAPS constraints are:

$$\sum_i \alpha_{ik} \times R_i \leq FGX_k \quad (A5.4)$$

Where:

FGX_k is the FAPS capacity on flowgate k

Dispatch Analogy

The auction constraints in equations A5.4 take the same form as transmission constraints in dispatch. They are not the *same* constraints as dispatch, since flowgate support generators are not included. Nevertheless, an *analogy* can be drawn as follows:

- the post-auction registered access, R, is analogous to the dispatched output, G;

262 Negative participation is ignored since this does not affect TFGX.

- the participation factors apply in both cases;
- the FAPS capacity is analogous to the flowgate capacity in dispatch;
- auction bids are analogous to dispatch offers; and
- auction clearing prices are analogous to LMPs from dispatch.

The value of this analogy is that some of the concepts that were demonstrated in relation to dispatch and pricing can also be applied to auction clearing. The dispatch is an analogy *only*. It is not being suggested that the auction would be cleared using NEMDE or other dispatch engine. A dedicated auction clearing engine would be developed.

There is an obvious difference between auction bid prices and dispatch offer prices. In the auction, the *higher* the bid price, the more likely the bid is to be *cleared*. In dispatch, the *lower* the offer price the more likely the generator is to be *dispatched*.

In fact, the relationship can be defined as follows:

Dispatch offer = RRP – auction bid price

The RRP outcome under the analogous dispatch can be assumed to be *fixed*: eg, by placing a very large generator and demand at the RRN.[263] If the RRN generator has a dispatch offer of $1,000/MWh, say, RRP is fixed at $1,000.

Auction Bids and Dispatch Offers

In general, a firm participant could:

- have some pre-auction registered access (Q);
- submit several bids in the form of price-quantity pairs (BP_j,BQ_j), j=1,2,...
- also submit several offers in the form of price-quantity pairs (OP_j,OQ_j), j=1,2,...

It is assumed that:

- the offers and bids are each ordered by *descending* price;
- offer prices are higher than bid prices: ie, the lowest offer price is higher than the highest bid price; and
- the *aggregate* offered quantity is permitted to be no higher than the pre-auction registered access.[264]

263 Recall that generation and demand place at the RRN does not change the transmission constraints.

264 The participant cannot offer what it does not have.

A dispatch offer consists of several price bands, each with a price/quantity pair. The analogous dispatch offer is constructed from the auction bids as follows:

- the first price band is for price -\$1,000[265] and for the *unoffered quantity*: the pre-auction holding minus the total offered quantity;
- the next price bands relate to the auction offers, with price and quantity ($RRP-OP_1$, OQ_1), ($RRP-OP_2$, OQ_2) etc; and
- the next price bands relate to the auction bids, with price and quantity ($RRP-BP_1$, BQ_1), ($RRP-BP_2$, BQ_2).

Since the auction offer and bid prices are in *descending* price order, the dispatch offer prices are in *ascending* price order. This is, of course, a requirement of a dispatch offer.

The offered quantity must be non-negative.

The construction of dispatch offers is illustrated by a couple of simple examples.

First, consider a firm participant with 100MW of pre-auction access who was bidding for an additional 50MW at \$100. The dispatch offer would be constructed as follows:

- the unoffered quantity, of 100MW, is offered into dispatch at -\$1,000; and
- the 50MW bid quantity is offered at RRP-bid = \$900.

Second consider a firm participant holding 200MW of pre-auction access and offering to sell 30MW of it at \$50.

The dispatch offer would be constructed as follows:

- the unoffered quantity, of 170MW, is offered into dispatch at -\$1,000; and
- the 30MW offered quantity is offered at RRP-offer = \$950.

Auction Clearing

The analogous dispatch problem is set up based on the FAPS constraints and auctions bids and offers following the steps discussed above:

- establish the FAPS constraints;
- receive bids and offers from auction participants;
- convert these into the analogous dispatch offers; and
- add a large generator and demand at the RRN.

---

265 Or any low number that ensures that it is dispatched.

An economic dispatch is then calculated for this dispatch problem, consisting of:

- dispatch targets, G_i; and
- local dispatch prices, LMP_i.

The auction clearing point is then defined as follows. The cleared auction quantity for each participant is determined by the formula:

$$AQ_i = G_i - Q_i$$

Where:

AQ_i is the quantity purchased in the auction (or sold, if AQ is negative)

Q_i is the pre-auction registered access

G_i is the dispatch target from the analogous dispatch

Auction clearing prices are determined by the formula:

$$P_i = RRP - LMP_i$$

Where

P_i is the clearing price for participant i

These cleared quantities and amounts satisfy the following principles:

1. clearing prices are consistent with bids and offer: equivalent to a "no regret dispatch" outcome;
2. post-auction the FAPS constraints are satisfied; and
3. subject to these constraints, the value of the auction is maximised.

Satisfaction of the no-regret dispatch principle is demonstrated in the next section. The FAPS constraints are satisfied because they are introduced directly into the dispatch analogy. The post-auction holdings are:

Post-auction holding = pre-auction holding + auction purchase

$$= Q + (G\text{-}Q) = G$$

Since the dispatch, G, satisfies the FAPS constraints, the post-auction holdings also satisfy the FAPS constraints.

Demonstrating that auction value is maximised is more complex. It relies on an understanding of the dual problem in linear programming and it is not presented here.

No-regret dispatch

It has previously been shown that LMPs are clearing prices that ensure no-regret dispatch: ie, a generator is dispatched if and only if the LMP is at or above its offer price.

Specifically, the quantity in dispatch offer price band is only dispatched if the LMP exceeds the associated dispatch offer price:

$$\text{LMP} \geq \text{dispatch offer price}$$

Since the first band, representing the unoffered quantity is priced at -$1000, this is always dispatched and so:

$$G \geq \text{unoffered quantity}$$

Therefore, the post-auction holding is never lower than the unoffered quantity.

Suppose that one of the dispatch offer bands, j, that relates to an auction *offer* is dispatched. This means that this offered quantity is *included* in G, the post-auction holding. Therefore, the associated auction offer has *not* been cleared.

Because the offer band is dispatched:

$$\text{LMP} \geq \text{dispatch offer price} = \text{RRP-OP}_j$$

Therefore:

$$\text{OP}_j \geq \text{RRP-LMP} = \text{P}_j$$

Therefore, the offer is *not* cleared if and only if the auction clearing price is *below* the auction offer price. That is consistent with auction principles.

Suppose now, instead, that one of the dispatch offer bands, j, that relates to an auction *bid* is dispatched. This means that this bid quantity is included in G, the post-auction holding. Therefore, the associated auction *bid* has been cleared.

Because the offer band is dispatched:

$$\text{LMP} \neq \text{dispatch offer price} = \text{RRP - BP}_j$$

$$\text{BP}_j \neq \text{RRP - LMP} = \text{P}_j$$

Therefore, the bid is cleared if and only if the auction clearing price is below the auction bid price. Again, that is consistent with auction principles.

It is seen then that the auction clearing satisfies the normal auction requirement that bids and offers are only cleared if the bid or offer price is below or above the clearing price, respectively.

E.3.2 Auction Settlement Surplus

Overview

It important for the integrity and sustainability of the short-term auction that auction *deficits* - where the amount payable to sellers exceeds the amount receivable from purchasers - do not arise. This section demonstrates that there will be no auction deficit so long as the pre-auction registered access complies with the FAPS constraints.

FAPS constraints

Suppose that each generator, i, has Q_i of access (at its local node) pre-auction and holds R_i of access post-auction. Thus if $R_i>Q_i$ it has purchased access at the auction and if $R_i<Q_i$ it has sold access. R_i cannot be less than zero, meaning that a generator cannot sell more into the auction than it held pre-auction.

Both pre-auction and post-auction holdings must be FAPS compliant. Therefore, for each flowgate k:

$$\sum_i(\alpha_{ik} \times Q_i) \leq FGX_k$$

$$\sum_i(\alpha_{ik} \times R_i) \leq FGX_k$$

Auction Payments

Now, the payment by each firm participant into the auction is the quantity purchased multiplied by the clearing price:

$$Pay\$_i = (R_i - Q_i) \times P_i \qquad (A5.5)$$

Recall that:

$$P_i = RRP - LMP_i$$

Where RRP and LMP are the values calculated in the analogous dispatch. The difference between RRP and LMP is based on the flowgate prices:

$$LMP_i = RRP - \sum_k \alpha_{ik} \times FGP_k$$

Therefore:

$$P_i = \sum_k (\alpha_{ik} \times FGP_k)$$

Where the FGPs are the flowgate prices determined in the analogous dispatch.

Substituting this expression into equation A5.5 gives:

$$Pay\$_i = (R_i - Q_i) \times \sum_k (\alpha_{ik} \times FGP_k) \qquad (A5.6)$$

Aggregate payments into the auction are determined by summing equation (A5.6) across all generators:

$$\text{Total Pay\$} = \sum_i \text{Pay\$}_i$$

$$= \sum_i \sum_k \{(R_i \times \alpha_{ik} \times FGP_k) - \sum_i \sum_k (Q_i \times \alpha_{ik} \times FGP_k)\}$$

$$= \sum_k FGP_k \times \{\sum_i \{(R_i \times \alpha_{ik}) - \sum_i (Q_i \times \alpha_{ik})\} \qquad \text{(A5.7)}$$

Because the post-auction holdings are FAPS compliant and the FGP is only positive if the relevant FAPS constraint is binding then either:

$$\sum_k (R_i \times \alpha_{ik}) < FGX_k \text{ and } FGP_k = 0$$

Or:

$$\sum_k (R_i \times \alpha_{ik}) = FGX_k \text{ and } FGP_k > 0$$

In both cases:

$$FGP_k \times \sum_i (R_i \times \alpha_{ik}) = FGP_k \times FGX_k$$

Therefore, substituting this into equation A5.7 gives:

$$\text{Total Pay\$} = \sum_k \{FGP_k \times [FGX_k - \sum_i (Q_i \times \alpha_{ik})]\} \qquad \text{(A5.8)}$$

Now, the pre-auction holdings, Q, also comply with the FAPS constraints:

$$\sum_i (Q_i \times \alpha_{ik}) \leq FGX_k$$

Therefore, the difference term inside the square brackets in equation (A5.8) is never negative and, since flowgate prices are also never negative:

$$\text{Total Pay\$} \geq 0$$

Entitlements Bundles

One way of understanding the algebra above is to think of nodal access as a bundle of *flowgate entitlements*. Registered access, A_i, at node *i*, provides a generator with a target firm entitlement amount, $A_i \, x \, a_{ik}$, on each flowgate *k*. At the auction, although generators bid and offer *nodal access*, it is effectively *flowgate entitlements* that are cleared in the auction, with clearing prices set equal to the flowgate prices appearing in the algebra above.

Thus, where a generator *i* buys $R_i - Q_i$ of access in the auction, it is essentially buying $(R_i - Q_i) \times \alpha_{ik}$ of flowgate entitlement on each flowgate k. It must pay the clearing price, FGP, for each entitlement, implying a total payment:

$$\text{Pay\$}_i = \sum_k [(R_i - Q_i) \times \alpha_{ik}] \text{ x } FGP_k$$

$$= (R_i - Q_i) \times \sum_k \alpha_{ik} \times FGP_k$$

Which is the same amount that it pays based on the nodal clearing prices, as derived in equation A5.6, above.

Since all entitlements on a flowgate are traded at a common price, FGP_k, the auction surplus arising from a particular flowgate *k*, is simply the flowgate price multiplied by the net aggregate amount of entitlements purchased on that flowgate which, in turn, is the difference between the total post-auction entitlement holdings and the total pre-auction entitlement holdings. The flowgate constraints ensure that, in both cases, the total holdings cannot exceed flowgate capacity.

E.3.3 Cross-regional settlement

Overview

The revenue from the short-term auction needs to be allocated between TNSPs. As noted in section 7.2.4 this will be based on flowgate tags.

The auction revenue can be attributed between flowgates, as discussed in this section. Flowgate revenue is based on the concept of flowgate entitlements being sold at a common price FGP_k, as discussed in the previous section. The revenue attributed to a flowgate is then paid to the TNSP who is tagged for that flowgate.

Flowgate Revenue

The auction revenue attributed to a flowgate is defined as:

$$\text{Flowgate\$}_k = FGP_k \times FGQ_k \quad \text{(A5.9)}$$

Where:

$\text{Flowgate\$}_k$ is the payment attributed to flowgate k

FGQ_k is the cleared quantity on flowgate k

The cleared quantity is defined by the equation:

$$FGQ_k = \sum_i (AQ_i \times \alpha_{ik}) \quad \text{(A5.10)}$$

Where AQ_i is the cleared quantity for participant i.

Auction Balance

For the auction to balance, it is necessary that:

$$\text{Auction revenue} = \sum_k \text{flowgate\$}_k$$

This is demonstrated below.

Recall from equation A5.7 that:

$$\text{Auction surplus\$} = \sum_k FGP_k \times \{\sum_i \{(R_i \times \alpha_{ik}) - \sum_i (Q_i \times \alpha_{ik})\}$$

$$= \sum_k FGP_k \times \{\sum_i \{(R_i - Q_i) \times \alpha_{ik})$$

$$= \sum_k FGP_k \times \{\sum_i \{(AQ_i) \times \alpha_{ik})$$

$$= \sum_k FGP_k \times FGQ_k \text{ (from A5.10)}$$

$$= \sum_k \text{Flowgate\$}_k \text{ (from A5.9)}$$

Positive Flowgate Payments

The flowgate payments are non-negative, as demonstrated below. If $FGP_k=0$ this is trivial.

If $FGP_k>0$ then the corresponding FAPS constraint must be binding, ie:

$$\sum i\ (R_i \times \alpha_{ik}) = FGX_k$$

Therefore:

$$FGQ_k = \sum_i (AQ_i \times \alpha_{ik})$$

$$= \sum_i \{(R_i \times \alpha_{ik}) - \sum_i (Q_i \times \alpha_{ik})$$

$$= FGX_k - \sum_i (Q_i \times \alpha_{ik})$$

≥ 0 because the pre-auction holdings are FAPS compliant

E.3.4 Tapering Capacity Release

Overview

Section 7.3.8 proposed that any spare short-term FAPS capacity should be released *progressively* over a sequence of short-term auctions. If there were a series of N auctions, then the release could be as follows:

- release 1/N of the capacity in the first auction;
- release 1/(N-1) of the remaining capacity in the second auction…and so on until...;
- release ½ of the remaining capacity in the penultimate auction; and
- release all the remaining capacity in the penultimate auction.

Therefore a general formulation is needed to ensure that 1/M of the remaining spare capacity is sold in the auction, for any positive integer M. This section describes such a formulation.

Scaling Up the Auction

In the revised auction formulation, the bids and offered quantities are scaled up by a factor M before they are processed into the dispatch offers as described previously.[266]

The analogous dispatch is then run as usual to calculate an economic dispatch, G, which complies with the FAPS constraints. However, the cleared quantity is now scaled back by M, ie:

$$AQ_i = (G_i - Q_i) / M$$

Spare Capacity

Define pre-auction spare FAPS capacity on a flowgate as:

$$SC^{pre}_k = FGX_k - \sum_i (Q_i \times \alpha_{ik})$$

The post-auction holding for generator i is:

$$R_i = Q_i + AQ_i = Q_i + (G_i - Q_i) / M$$

$$= G_i / M + (1 - 1 / M) x Q_i$$

The post-auction spare capacity on flowgate k is therefore:

$$SC^{post}_k = FGX_k - \sum_i (R_i \times \alpha_{ik})$$

$$= FGX_k - \sum_i \{[(G_i / M) + (1 - 1 / M) \times Q_i] \times \alpha_{ik}\}$$

$$= FGX_k - \sum_i (G_i \times \alpha_{ik}) / M - (1 - 1 / M) \sum_i (Q_i \times \alpha_{ik})$$

$$\geq FGX_k - FGX_k / M + (1 - 1 / M) \sum_i (Q_i \times \alpha_{ik})$$

$$= (1- 1 / M) \times (FGX_k - \sum_i (Q_i \times \alpha_{ik}))$$

$$= (1 - 1 / M) \times SC^{pre}_k$$

Therefore, no more than 1/M of the pre-auction spare capacity has been sold. In fact, where *less* than 1/M was sold, the corresponding dispatch constraint was not binding: ie the FGP was zero. So, 1/M of the pre-auction spare capacity has been released and is *available* to be sold, where it has some value.

266 Note that this could lead to the unoffered quantity being negative. This does not matter at a conceptual level. If this created a problem for the dispatch process, a local demand could be introduced to replace the negative unoffered quantity.

A simple example can clarify this. Suppose that:

- a particular flowgate has FAPS capacity of 1,000MW;
- 150MW is spare pre-auction;
- M=10, so that 1/10 of the spare capacity (ie 15MW) is to be released at the auction;
- the flowgate is radial with two participating generators, X and Y; and
- generator X bids for 8MW at $20 and the Y bids for 10MW at $15.

In the auction the bids are first scaled up by 10, so:

- generator X bids for 80MW at $20; and
- generator Y bids for 100MW at $15.

The auction is run with *all* the spare capacity released. This means that X, nominally, buys 80MW and Y buys 70MW. However, the cleared quantities are one tenth of the nominal quantities, ie:

- 8MW is sold to generator X;
- 7MW is sold to generator Y; and
- 15MW is sold in total: ie, one tenth of the pre-auction spare capacity.

Complies with Auction Principles

The scaled auction, as well as releasing 1/M of the spare capacity, must comply with the other auction principles:

- cleared quantities are consistent with bids and offers;
- post-auction holdings are FAPS compliant; and
- subject to these constraints the auction value is maximised.

Clearly, the scaled auction complies with these constraints, since it is just like any other auction, just with different bids and offers. In particular, cleared quantities are consistent with bids and offers:

> Nominal clearing quantity ≤ nominal bid/offer = M x actual bid/offer quantity

The *true* clearing quantity is just 1/Mth of the nominal amount:

> True clearing quantity = 1/M x nominal clearing quantity ≤ actual bid/offer quantity

The no-regret principle is unaffected by the scaling. So a bid or offer still clears only if the clearing price is below or above the bid or offer price, respectively.

The post-auction holding is FAPS compliant because:

Post-auction spare capacity $\geq$ (1-1/M) x pre-auction spare capacity $\geq$ 0 and M$\geq$1

As before, optimality is more complex and is not demonstrated here.

F TNSP Regulation

F.1 TUOS Neutrality

F.1.1 Overview

Section 8.2.2 explains how a timing mismatch could occur within a regulatory period such that there is a difference between:

- access revenue; and
- carrying costs of assets delivering firm access.

This could occur even when access pricing is accurate: that is to say, where the access price exactly matches the true expansion costs associated with delivering the access.

Revenue regulation would allow this mismatch to be passed through to TUOS. Importantly, this does not mean that TUOS customers bear this mismatch risk instead. This is demonstrated in this section.

F.1.2 Present Costs

Present costs and value incorporate all future expenditure or revenue, discounted by the relevant discount rate. In the context of TNSP regulation, the TNSP regulated WACC is used as the discount rate.

The access price is set at the expected present cost of the incremental expansion costs associated with providing access. If the price is accurate then:

Access price = present cost(incremental expansion costs)

At the time of access procurement, a payment profile is established such that the present value of the access payments equals the access price:

Access price = present value(access payments)

Finally, the capital cost of expansion can be amortised into annual carrying costs, taking the form:

Carrying cost = WACC x DRC + D

Where:

WACC = regulated WACC for TNSP

DRC = depreciated replacement cost

D = annual depreciation

For simplicity it is assumed that replacement costs do not vary: ie, the replacement cost is equal to the historical cost. In this case, the present cost of the annual carrying costs for a particular network asset equals the capital cost:

Capital cost = present cost(carrying costs)

Aggregating this equation across all assets associated with the incremental expansion required to deliver the access:

Present cost(incremental expansion costs) = present cost(incremental carrying costs)

Therefore, the following three present costs are equal:

- access payments;
- incremental expansion costs; and
- incremental carrying costs.

F.1.3 Regulatory Periods

These present costs can be divided between the current regulatory period and future regulatory periods. All of the components continue to be defined in present cost terms, so that the sum of the components equals the present cost:

Present cost(access payments) = $AP^c + AP^f$

Present cost(incremental carrying costs) = $CC^c + CC^f$

Where:

AP^c = access payments in current period

AP^f = access payments in future periods

CC^c = carrying costs in current period

CC^f = carrying costs in future period

Because the present costs are equal:

$$AP^c + AP^f = CC^c + CC^f \quad \text{(A6.1)}$$

Now consider the situation where some long-term firm access is issued by a TNSP within a regulatory period that was not forecast at the time of the previous reset. The revenue regulation rules are:

- the extra access payments in the current regulatory period (AP^c) are retained by the TNSP;

- the extra carrying costs in the current regulatory period (CC^c) are borne by the TNSP;
- the difference between the two amounts (CC^c-AP^c) above is added to the AARR in the following regulatory period;
- new assets developed in the current regulatory period are rolled into the asset base at the period end, such that all remaining carrying costs (CC^f) are recoverable from the TNSP AARR; and
- access payments outstanding at the end of the current regulatory period (AP^f) are allowed for when setting the TUOS revenue cap for future regulatory periods.

Therefore, for future regulatory periods, the new access adds to the AARR:

$$\text{Increase in AARR} = CC^f + (CC^c - AP^c)$$

There is also an increase in access revenue of AP^f. Therefore, the increase in the TUOS revenue cap is:

$$\text{Increase in TUOS revenue cap} = \text{increase in AARR} - \text{increase in access revenue}$$

$$= CC^f + (CC^c - AP^c) - AP^f$$

$$= (CC^c + CC^f) - (AP^c + AP^f)$$

$$= 0 \text{ (from equation A6.1)}$$

Therefore, there is no net impact on the TUOS revenue cap from the new access.

F.2 Contingent Auction in a reliability RIT-T

F.2.1 Overview

Section 8.2 outlines a mechanism through which firm access could be issued through a contingent auction held as part of the RIT-T for a reliability expansion. This section provides more details of this mechanism.

F.2.2 Access Pricing

A contingent auction would become relevant where one or more of the reliability expansion options being considered in the RIT-T creates some additional reliability access. This will be generally reflected in a reduction in the associated (firm) access price. That is:

$$AP^{post} < AP^{pre}$$

Where:

AP^{pre} is the access price based on the existing network topology

AP^{post} is the access price when the expansion option is added to the existing network topology

Such a price impact will typically arise from those expansions that add capacity on the *generation-side* network: ie, between some generator nodes and the RRN. These expansions will, by definition, create some RA and so should be subject to a contingent auction process.

On the other hand, demand-side expansions – for example addition of transformer capacity at a distribution connection point – that do not affect access prices would not create RA and so do not require a contingent auction. For these expansions, the current RIT-T process would be unchanged.

F.2.3 Auction Bids

Where this price impact is expected, for a particular expansion option, bids for long-term firm access would be invited from the associated generators (obviously, not all generators will see such a price impact).

A generator can bid at any price (or not at all). Clearly, it is unlikely to bid at, or above, the current access price, since it is able to purchase firm access at that price through the normal access procurement process. On the other hand, assuming that the particular expansion option goes ahead, there is a benefit to the TNSP of accepting a bid that is higher than AP^{post}, since the access revenue then more than covers the incremental expansion cost: assuming, as always, that access pricing is accurate.

For example, suppose that:

AP^{pre} = \$160m

AP^{post} = \$90m

Bid = \$120m

If the expansion proceeds, there is \$30m (\$120m - \$90m) benefit to the TNSP of accepting the bid.

F.2.4 Contingent Benefits

The benefit arising from accepting the bid is contingent on the particular expansion option proceeding. The role of the RIT-T is to identify the preferred expansion. The RIT-T is required to evaluate all market costs and benefits associated with an expansion option and this should include the contingent benefit discussed above. An example of such a revised RIT-T process is presented in Table F.1 below. There are two options being compared in the RIT-T. Each satisfactorily addresses the requirement to maintain reliability standards. The RIT-T requires that the option with the lower net cost (or higher net benefit) is selected.

Table F.1 RIT-T example

Expansion Option	Option 1	Option 2
Associated Generator	Gen A	Gen B
Expansion Cost	$100m	$150m
AP^{pre}	$160m	$280m
AP^{post}	$90m	$130m
Gen Bid	$120m	$220m
TNSP Contingent Benefit	$30m	$90m
Net Cost of Option	$70m	$60m

It is seen that, once the contingent benefit is included, Option 2 has the lower net cost (of $60m = $150m - $90m) and so would be selected for development. If the contingent auction had *not* been run, and so the contingent benefit was not included, Option 1 would instead have been selected, since it has the lower expansion cost. Therefore, the inclusion of the contingent auction improves the efficiency of transmission planning.

F.2.5 Auction Process

The auction process would operate within the RIT-T as follows:

1. a TNSP identifies a set of expansion options as usual;
2. a TNSP publishes these options and invites generators to make bids for firm access. Each bid would be associated with a particular expansion option;
3. the TNSP calculates the contingent benefit for each option and includes this within the RIT-T analysis;
4. the TNSP identifies and selects the preferred option;
5. bids associated with that option are cleared. Other bids lapse; and
6. access procurement/issuance completes in the usual way: ie, a payment profile is established and included in a payment deed; firm access certificates are issued.

Generators would need to be informed about the impact of the expansion options on access prices: ie, the values of AP^{post}. That could be done, for example, by the TNSP including the expansion options in different versions of the access pricing model.

The framework for the bidding and auction would have to lock generators in. Otherwise, a generator might change its bid once it learned of the preferred option and this would, in turn, undermine the case for the preferred option.

The example has only one generator bidding on each option. In practice, to get an efficient outcome, there would need to be competing bids on each option. In this case, only the highest value bid or bids would be cleared for the preferred option.

G Glossary

Table G.1 OFA Technical Glossary

Defined Term	Meaning
Access	*Network access*.
(Network) access	The MW amount for which a generator or directed interconnector is paid the difference between RRP and LMP in AEMO settlement.
Access charge	The amount payable, in total, to a TNSP for procured firm access.
Access firmness	The average level of access provided (possibly weighted by the congestion price) compared to the registered access.
Access issuance	A process through which a TNSP provides new firm access.
Access payment	An instalment of the access charge.
Access price	A regulated price for some specified firm access which reflects the long-run incremental cost to a TNSP of providing that access
Access procurement	A process through which a market participant purchases firm access from a TNSP or another firm participant.
(Firm) access registration	A record of an existing or prospective *firm access service* provided by a TNSP, or TNSPs, to a market participant, specifying all of the *service parameters* of that service.
Access request	A formal request for new firm access, made to a TNSP, with specified service parameters.
Access settlement	A new AEMO settlement process in the OFA model which makes payments based on the difference between access and dispatch.
Access unit	The generator entity that participates in access settlement. A dispatchable unit or a group of dispatchable units whose output is measured by a common revenue meter.
Adjusted cost	The present cost of the adjusted network development scenario.
Adjusted network development scenario	A set of stylised expansions that are forecast to be required in order to maintain FAPS under an adjusted scenario.
Annual benchmark	An annual dollar amount, set by the AER, which is used in the operational incentive scheme. It is based on the estimated operating cost and capped shortfall cost of an efficient TNSP.
Auxiliary load	Load that is related to power station operation and which is deducted from generation output when calculating sent-out output.
Availability	For a conventional generator, the offered availability; for an intermittent generator, the unconstrained intermittent generation forecast.

Defined Term	Meaning
Baseline cost	The present cost of the baseline network development scenario.
Baseline forecast	A sequence of annual FAPS snapshots, together with forecast registered access, over a number of consecutive years, used for access pricing.
Baseline network development scenario	A set of stylised expansions that are forecast to be required in order to maintain FAPS under a baseline scenario.
Capped shortfall cost	The shortfall cost after the nested caps specified in the incentive scheme have been applied. Forms the basis for the TNSP penalties and rewards payable under the incentive scheme.
Completed (access) request	An access request that leads to firm access being issued.
Competing (access requests)	A property of two access requests, where their access prices materially depend upon the order in which they are processed.
Congested flowgate	A *flowgate* whose capacity is fully utilised in dispatch and which is causing dispatch to be constrained.
Congestion	A dispatch condition in which one or more flowgates are congested.
Congestion price	The component of the difference between the region price and the local price that is unrelated to losses.
Congestion rent	The component of IRSR that arises from congestion on inter-regional flowgates and which is allocated to DICs as part of access settlement.
Constrained off	(For a *generator)* dispatched below its *preferred output*.
Constrained on	(For a *generator*) dispatched above its *preferred output*.
(NEMDE) constraint	A potential constraint on dispatch which is included in NEMDE.
Constraint equation	A linear inequality representing a NEMDE constraint.
Counterprice flow	A flow on an interconnector, which is directed towards the RRN with the lower RRP.
Deep connection cost	The immediate (but not future) incremental costs to a TNSP associated with providing additional firm access: ie, only including those costs that must be incurred prior to access commencement.
Directed interconnector	An *interconnector* in a specified direction: ie, northerly or southerly. A conceptual, inter-regional entity that participates in AEMO settlement.
Dispatch access	The right to be dispatched in NEM dispatch at a specified MW level in accordance with a dispatch offer and paid the locational marginal price on dispatched output.
Dispatchable unit	Either an individual generating unit or logically grouped generating units that are connected to the same node.

Defined Term	Meaning
Economic expansion	A network expansion that is not required for maintaining reliability standards.
Effective access	The level of network access that a generator receives through the combination of regional settlement and access settlement.
Effective flowgate capacity	The *flowgate capacity* plus the *flowgate support*.
Embedded generator	A distribution-connected *generator*.
Entitlement	A flowgate entitlement.
Entitlement allocation	The determination of entitlements to flowgate access generators and DICs such that they sum to the effective flowgate capacity.
Existing capacity	The capacity of generators at the time of OFA commencement, on which allocation of TA is based.
Exporting region	The region from which a *directed interconnector* withdraws power.
FAPS capacity	The level of effective flowgate capacity under *FAPS conditions*.
FAPS conditions	The network and market conditions to which the *Firm Access Planning Standard* refers, for which the TNSP must plan to provide *target flowgate capacity*.
FAPS demand	The regional demand under FAPS conditions.
FAPS requirement	The requirement that FAPS capacity is no lower than TFGX on congested flowgates.
Firm access (service)	A transmission service provided by TNSPs to generators and *directed interconnectors*.
Firm access certificate	Proof of firm access purchase, issued by a TNSP to the purchaser.
Firm access operating standard	The operating component of the firm access standard.
Firm access planning standard	The planning component of the firm access standard.
Firm access register	A database, maintained by AEMO, of firm access registrations.
Firm access standard	The service standard for firm access, which is the lowest level of service quality that the TNSP is permitted to provide.
Firm generator	A generator with *registered access* equal to its *capacity*.
Firm generation dispatch	A simultaneous dispatch of all generators and DICs at their registered access level under FAPS conditions. A balancing load or generator is added at each RRN.

Defined Term	Meaning
Firm interconnector right	The right to receive a portion of the settlement payments made to a specified DIC.
Firm participant	A firm generator or firm DIC.
Flowgate	A point of potential congestion on the transmission network; the notional location on a transmission network represented in NEMDE by a transmission constraint.
Flowgate access generator	A generator with positive participation in a particular flowgate, meaning that its output exacerbates congestion on the flowgate.
Flowgate capacity	The maximum aggregate usage of a flowgate allowed in dispatch. The RHS of the corresponding NEMDE transmission constraint.
Flowgate entitlement	The amount of a flowgate to which a generator (or DIC) is entitled. When usage exceeds the entitlement, the generator must make payments into access settlement.
Flowgate price	The marginal value of flowgate capacity in dispatch: the amount by which the total cost of dispatch would increase if flowgate capacity were reduced by 1MW; calculated in NEMDE as the dual value of the corresponding transmission constraint. The rate at which a generator pays into access settlement when its usage exceeds its entitlement.
Flowgate support	The aggregate, absolute flowgate usage of flowgate support generators.
Flowgate support generator	(With respect to a flowgate) a generator with a participation factor less than zero. Its output relieves congestion on a flowgate.
Flowgate tag	An attribute of a flowgate which defines which TNSP or TNSPs are responsible for that flowgate under FAS.
Flowgate tagging	The determination of flowgate tags.
Flowgate usage	The amount of a generator's output notionally flowing through the flowgate; the product of the generator's output and its flowgate participation.
Generator	Either an access unit or the generating company responsible for the access unit, depending upon the context.
Generator node	The transmission or distribution node at which a generator, connects to the shared transmission network.
Importing region	The region into which a directed interconnector injects power.
(Operational or FAOS) incentive scheme	A requirement of FAOS, under which a TNSP is incentivised to efficiently manage the market cost of flowgate shortfalls.
Incremental usage	The increased usage on a flowgate associated with an access request.
Interconnector	A notional entity that is dispatched by NEMDE to transfer power between two RRNs, across a regulated interconnector.
Inter-regional	Network access provided to a directed interconnector, from the RRN in

Defined Term	Meaning
access	the exporting region to the RRN in the importing region.
Inter-regional flowgate	A flowgate in which interconnectors, and possibly also generators, participate.
Inter-regional hedge	A security which pays out an amount proportional to the inter-regional price difference in a settlement period, used by market participants to hedge inter-regional price risk.
Inter-regional price difference	The difference in RRP between two neighbouring regions.
Inter-regional settlement residue	The surplus from regional settlements that is attributed to DICs. The sum of the congestion rent and the losses residue.
Intra-regional access	Network access provided to a generator, from its generator node to the RRN in its local region.
Intra-regional flowgate	A flowgate in which only generators, not interconnectors, participate.
Lapsed (access) request	An access request which is not completed.
Local node	The location at which a generator is connected to the shared network.
Local price	The amount that a generator is paid under OFA for a marginal increase in dispatch output. For flowgate access generators, under normal conditions, this equals the locational marginal price.
Local region	(Of a generator) the region in which the generator is connected.
Locational marginal price	The marginal value that a generator at a node provides to economic dispatch.
Long-run	In the context of access procurement, the period in which flowgate expansion is likely to be feasible; beyond the transmission expansion lead time.
Long-run decremental cost	The immediate and future incremental savings to a TNSP associated with no longer having to provide some firm access.
Long-run incremental cost	The immediate and future incremental costs to a TNSP associated with providing additional firm access.
Long-run marginal cost	The long-run incremental cost calculated assuming no lumpiness of transmission expansion and no spare transmission capacity.
Losses residue	The component of inter-regional settlement residue that is not related to congestion and which is paid to TNSPs.
Maximum usage	The maximum possible aggregate usage of a flowgate. If this is less than the flowgate capacity, then the flowgate cannot become congested.
Meshedness	Attribute of a network element which reflects the number of alternative paths between the end nodes. If the meshedness equals one then the element is radial and removing it would divide the network into two

Defined Term	Meaning
	islands.
NEMDE	National Electricity Market Dispatch Engine: the computer application through which AEMO calculates dispatch targets for scheduled plant in the NEM and calculates dispatch prices.
Network element	An entity in a network model that represents a transmission line or network transformer.
Network topology	The specification of a transmission network in a network model. It contains the electrical characteristics (eg, admittance and rating) but not geographical characteristics (eg, line length or location).
Node	A local node or a regional reference node.
Non-firm access	The access received by a non-firm generator.
Non-firm generator	A generator with no *registered access*.
No-regret dispatch	The situation where the extra payment to a generator from being dispatched equals or exceeds the associated dispatch cost.
Non-scheduled generator	A generator that is not dispatched by AEMO but instead chooses its level of output. It does not participate in access settlement and continues to be paid the RRP.
Queueing policy	A policy which determines the order in which access requests are processed.
Parallel interconnectors	Interconnectors which flow power between the same two RRNs.
Part-firm generator	A generator with some registered access that is less than its capacity.
(Flowgate) participation (factor)	The proportion of a *generator's* output that uses a *flowgate*; the coefficient applied to that generator's dispatch variable in the left hand side of the corresponding NEMDE constraint equation.
Payment profile	The schedule of access payments associated with an access charge.
Power system stability	The ability of the power system to maintain voltage, frequency and phase angles within specified limits under steady-state and post-contingency conditions.
Preferred output	The quantity of a generator's availability that is offered at or below the RRP.
Radial (constraint or flowgate)	A flowgate in which all participation factors are either unity or zero. Typically, a thermal limit on a radial element: one whose removal would split the network into two islands.
Rated capacity	The maximum determined export capacity of an access unit.
Regional settlement	The existing settlement arrangements in which generators are paid the loss-adjusted regional price for their output.
Registered access	For a generator, the nominal amount of access specified in a firm access registration. For an FIR holder, the nominal amount of the FIR. For a

Defined Term	Meaning
(amount)	DIC, the aggregate amount of FIRs registered on that DIC.
Reliability access	Access provided to non-firm generators as a result of a TNSP expanding transmission to meet a demand-side reliability standard.
Reliability expansion	A network expansion that is required in order to maintain reliability standards.
Reliability generator	A generator who receives reliability access.
Reliability standard	The minimum service requirement for TNSP supply to consumers.
Regional price	The regional reference price.
Regional original price	The locational marginal price at the regional reference node.
Regional reference node	A specified node in each region which is used in setting the regional reference price.
Regional reference price	The price paid to a dispatched generator in regional settlement. This is normally set equal to the regional original price, but is administered under certain conditions.
Remote region	A region other than the local region.
Renewal request	An access request that it made by a generator to replace some existing firm access that is expiring and which holds renewal rights. Renewal requests are priced at LRDC rather than LRIC.
Renewal right	An attribute of firm access where its renewal is anticipated in the baseline forecast.
Security adjustment	An adjustment to the estimated spare capacity on a network element, to reflect the higher flow on the element that might occur following a contingency.
Secondary trade	A direct or indirect trade and transfer of firm access between two market participants.
Sellback	The process under which a TNSP cancels a firm access registration and the associated generator receives a corresponding payment from the TNSP, based on the current LRDC of that firm access.
Sellback right	The right of a generator to request a sellback to a TNSP, which the TNSP is obliged to undertake.
Service parameters	Values contained in a firm access registration specifying the service provider, the service recipient, the access amount, the term, the location, and the generator (for intra-regional access) or directed interconnector (for inter-regional access).
Settlement period	The time interval for which access settlement variables are calculated. It is proposed that this equals a trading interval (30 minute period).
Settlement residue	The auction through which AEMO sells SRA rights.

Defined Term	Meaning
auction	
(Flowgate) shortfall	Any amount by which effective flowgate capacity is below target flowgate capacity.
Shortfall cost	The flowgate shortfall multiplied by the flowgate price.
Short-run	In the context of access procurement, the period in which flowgate expansion is likely to be infeasible; within the transmission expansion lead time.
Spare (network) capacity	The difference between TFGX and FAPS capacity.
SRA right	The right to receive a specified proportion of the inter-regional settlement residue for a specified directed interconnector.
Stability flowgate	A flowgate that is not a thermal flowgate.
Stylised expansion	A network expansion or replacement which is defined within the access pricing model.
Super-firm generator	A generator with registered access greater than its capacity.
Target access	The amount of access that would be allocated to a generator or DIC were there sufficient flowgate capacity. It is calculated as the sum of the target firm access and the target non-firm access.
Target firm access	(For a generator) the lower of the generator's *registered access* and *capacity*.
Target firm entitlement	(For a generator on a congested flowgate) the product of the target firm access and the participation factor.
Target flowgate capacity	The amount of flowgate capacity required to provide all generators and DICs with their target firm access. It equals the aggregate of the target firm entitlements. The amount of effective flowgate capacity that a TNSP must provide under FAPS conditions.
Target non-firm access	(For a generator) the difference between availability and registered access, when the former takes a higher value.
Target non-firm entitlement	(For a generator on a congested flowgate) the product of the target non-firm access and the participation factor.
Thermal flowgate	A flowgate that relates to a constraint designed to ensure that a thermal limit is not exceeded.
Thermal limit	The maximum capacity at which a transmission element can securely or safely operate without physical damage or risks caused by overheating.
Transitional access	A level of firm access service that is allocated to existing generators at the commencement of the optional firm access regime.
Transmission constraint	A constraint included in NEMDE that arises as a result of limitations on a shared transmission or distribution network and for which a constrained

Defined Term	Meaning
	generator is not compensated under current arrangements.
Weighted average cost of capital	The funding cost of a generator or TNSP. A blend of the costs of equity and debt.